



Finance Implications of the Great Recession

Citation

Zwick, Eric Meinberg. 2014. Finance Implications of the Great Recession. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12274329>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Finance Implications of the Great Recession

A dissertation presented

by

Eric Meinberg Zwick

to

The Department of Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

May 2014

© 2014 Eric Meinberg Zwick

All rights reserved.

Dissertation Advisors:

Professor David Scharfstein

Professor Andrei Shleifer

Author:

Eric Meinberg Zwick

Finance Implications of the Great Recession

Abstract

Macroeconomic events in the United States during the last ten years—the housing bubble, the financial crisis and the subsequent, deep recession—brought several puzzles to the attention of economists and policymakers. Why were there such large price booms and busts in places like Las Vegas and Phoenix, where land was readily available and construction markets were very active? Why are economies so slow to recover from recessions that coincide with financial crises? Can policymakers use fiscal stimulus to increase output and accelerate economic recovery after a recession?

My dissertation research contributes to how we think about these questions. Using ideas from finance and behavioral economics and a variety of methods—economic theory, new data, natural experiments and anecdotal exploration—I document several surprising facts. In the case of housing bubbles, introducing speculation into the standard supply-and-demand framework can reverse the conventional wisdom, so that supply can increase price volatility and exacerbate booms. In the case of policy responses to recessions, forcing banks to close can speed up local lending and financial frictions can amplify the effects of fiscal stimulus.

Contents

Abstract	iii
Acknowledgments	vi
Introduction	1
1 Do Financial Frictions Amplify Fiscal Policy? Evidence from Business Investment Stimulus	4
1.1 Introduction	4
1.2 Hypothesis Development	11
1.3 Business Tax Data	16
1.4 The Effect of Bonus Depreciation on Investment	21
1.5 Explaining the Large Response with Financial Frictions	34
1.5.1 Heterogeneous Responses by Ex Ante Financial Constraints	35
1.5.2 Heterogeneous Responses by Tax Position	38
1.5.3 Discount Rates and the Shadow Cost of Funds	42
1.6 Substitution Margins and External Finance	45
1.7 Conclusion	47
2 Arrested Development: Theory and Evidence of Supply-Side Speculation in the Housing Market	49
2.1 Introduction	49
2.2 Stylized Facts of the U.S. Housing Boom and Bust	56
2.2.1 The Cross-Section of Cities	56
2.2.2 The Central Importance of Land Prices	59
2.2.3 Land Market Speculation by Homebuilders	62
2.3 A Housing Market with Homeowners and Developers	64
2.4 Supply-Side Speculation	70
2.4.1 Land Speculation and Dispersed Homeownership	71
2.4.2 Belief Aggregation	73
2.5 The Cross-Section of City Experiences During the Boom	76
2.6 Variation in House Price Booms Within Cities	83

2.6.1	Location	83
2.6.2	Structure Type	84
2.7	Conclusion	85
3	Regulators vs. Zombies: Loss Overhang and Lending in a Long Slump	87
3.1	Introduction	87
3.2	Related Literature	90
3.3	Overview of U.S. Banking Regulation	91
3.3.1	Prompt Corrective Action	91
3.3.2	FDIC Resolutions	92
3.4	Data	93
3.5	Bank Failures in the Great Slump	94
3.6	Resurrected Lending following Resolution	100
3.6.1	Results from the Basic Regression	100
3.6.2	Alternative Explanations and Secondary Tests	105
3.7	Reduced Growth in Zombie-Afflicted Counties	110
3.7.1	FDIC Constraints	112
3.7.2	Local Economic Effects of Zombie Banks	114
3.8	Conclusion	118
	References	119
	Appendix A Appendix to Chapter 1	127
A.1	Investment with Adjustment Costs and a Borrowing Constraint	127
A.1.1	General Setup	127
A.1.2	Testable Hypotheses	131
A.1.3	Empirical Moments for Calibration	134
A.2	Legislative Background	135
	Appendix B Appendix to Chapter 2	146

Acknowledgments

I thank Raj Chetty, David Laibson, Josh Lerner, David Scharfstein and Andrei Shleifer for extensive advice and support. I thank the Harvard Business School Doctoral Office for financial support.

For help with Chapter 1, we thank Jediphi Cabal, Gary Chamberlain, George Contos, Ian Dew-Becker, Fritz Foley, Paul Goldsmith-Pinkham, Robin Greenwood, Sam Hanson, Ron Hodge, John Kitchen, Pat Langetieg, Day Manoli, Isaac Sorkin, Larry Summers, Adi Sunderam, Nick Turner, Danny Yagan and seminar and conference participants at Dartmouth, the Federal Reserve Board, Harvard, the IRS, Northwestern, Oxford, the US Treasury, the University of Chicago, the University of Texas, Washington University in Saint Louis, and Yale for comments, ideas, and help with data. We are grateful to our colleagues in the US Treasury Office of Tax Analysis and the IRS Office of Research, Analysis and Statistics—especially Curtis Carlson, John Guyton, Barry Johnson, Jay Mackie, Rosemary Marcuss and Mark Mazur—for making this work possible. The views expressed here are ours and do not necessarily reflect those of the US Treasury Office of Tax Analysis, nor the IRS Office of Research, Analysis and Statistics.

For help with Chapter 2, we thank Jediphi Cabal, John Campbell, Edward Glaeser, David Laibson, and Andrei Shleifer for outstanding advice and Robin Greenwood, Sam Hanson, Alp Simsek, Amir Sufi, Adi Sunderam, and Jeremy Stein for helpful comments. We also thank Harry Lourimore, Joe Restrepo, Hubble Smith, Jon Wardlaw, Anna Wharton, and CoStar employees for enlightening conversations and data. Nathanson thanks the NSF Graduate Research Fellowship Program, the Bradley Foundation, and the Becker Friedman Institute at the University of Chicago for financial support.

For help with Chapter 3, I thank Effi Benmelech, Jediphi Cabal, John Campbell, Shawn Cole, Paul Gompers, Robin Greenwood, Sam Hanson, Victoria Ivashina, Ben Iverson, Bill Kerr, David Laibson, Chris Malloy, Brock Mendel, Ramana Nanda, Jack Reidhill, David Scharfstein, Andrei Shleifer, Jeremy Stein, and Adi Sunderam for comments and ideas.

I thank Nickolaus Groh, Emma Trucks, and especially Jordan Padnuk for putting up

with my nonsense and supporting me throughout this process. I thank David Kushner and Joe Presbrey for unconditional love.

To Jediphi Cabal, my sisters, and especially my parents

Introduction

Macroeconomic events in the United States during the last ten years—the housing bubble, the financial crisis and the subsequent, deep recession—brought several puzzles to the attention of economists and policymakers. Why were there such large price booms and busts in places like Las Vegas and Phoenix, where land was readily available and construction markets were very active? Why are economies so slow to recover from recessions that coincide with financial crises? Can policymakers use fiscal stimulus to increase output and accelerate economic recovery after a recession? My dissertation research contributes to how we think about these questions. I apply insights from finance and behavioral economics while using a variety of methods: new data, natural experiments, anecdotal exploration and rigorous theory.

Two questions that have long intrigued economists and policymakers are whether tax policy affects firm behavior and whether credit market imperfections force firms to forgo profitable projects. Past research has struggled to find compelling natural experiments to address these questions. Experiments are necessary in the case of tax policy because taxes often change based on how the economy is performing, and in the case of credit market imperfections because firms will tend to borrow less when their prospects look worse. The first paper of my dissertation develops an experiment based on the idea that tax policy can affect different firms in different ways, and we use tax changes to create an “as-if” experiment in the amount of cash firms need to pursue their desired projects. By focusing on policies that were an important part of the 2009 U.S. stimulus bill, we also provide new evidence on how and when stimulus policy can be effective.

An important innovation of our study is that we work with a dataset which includes more than one hundred thousand public and private companies in the United States, built from a database of two million corporate tax returns filed between 1993 and 2010. Most studies of tax policy and investment behavior rely upon financial accounting data for public companies or aggregate investment data. These sources are limiting in several ways. First, public firms have broader access to capital markets than smaller, private firms, so the sample of public firms does not offer the ideal setting for documenting credit constraints. Second, with aggregate data it is difficult to isolate potential confounds since these policies do not happen randomly, and impossible to uncover the mechanism of any observed response. In contrast, our data include a granular account of investment activity; a precise tax account; a large sample of small, private firms; and measures of capital structure, payouts and employment. As a result, we are able to document how firms respond to tax policy at an unprecedented level of detail.

We present three findings. First, bonus depreciation raised investment 17.3 percent between 2001 and 2004 and 29.5 percent between 2008 and 2010. Second, financially constrained firms respond more than unconstrained firms. Third, firms respond strongly when the policy generates immediate cash flows but not when benefits only come in the future. The implied discount rate firms apply to future cash flows is too high to match a frictionless model and cannot be explained entirely by costly finance.

The second paper of my dissertation applies insights from behavioral finance to explain why cities with ample land experienced dramatic house price booms and busts in the early 2000s. This is a puzzle that economists have struggled to solve, but which is critical because these relatively unconstrained areas—including Arizona, inland California, Nevada and Florida—became the epicenter of the subprime and foreclosure crises.

The paper is methodologically diverse, combining national, subnational and within-city statistical analysis with rich anecdotal detail about Las Vegas, Phoenix and the behavior of public homebuilders culled from the footnotes of their annual reports. Our theory focuses on speculation by investors and homebuilders on the supply side of the market. Speculation

is easier in the land market than in the housing market due to rental frictions. Therefore, speculation amplifies house price booms the most in cities with ample undeveloped land. This observation reverses the standard intuition that cities where construction is easier experience smaller house price booms. It also explains why the largest house price booms in the United States between 2000 and 2006 occurred in areas with elastic housing supply. These episodes are most likely to occur in elastic cities approaching a long-run development constraint.

The third paper of my dissertation studies bank failures and how regulatory intervention affects local lending activity. It is motivated by an old question in economics: why are economic recoveries following financial crises so sluggish? I present micro-level evidence that, in the recovery from the U.S. Great Recession, loss overhang in the banking sector restricted lending and slowed growth. Zombie banks suffer from a debt overhang problem caused by unrealized losses on past loans. To deter regulatory action, zombies restrict new lending in healthy categories, prop up lending in unhealthy categories, and overallocate to safe, liquid assets. FDIC-induced failures allow zombies to hive off bad loans and as a result lending resumes post resolution. In the slump that began in the United States in 2007, limited FDIC liquidity and manpower prevented it from a timely reboot of all zombie balance sheets. As a consequence, counties afflicted with unhealed zombies displayed a slower recovery in employment, even in tradable goods industries less subject to local demand conditions.

Taken together, the research in this dissertation demonstrates how insights from finance can help answer important macroeconomic and policy questions.

Chapter 1

Do Financial Frictions Amplify Fiscal Policy? Evidence from Business Investment Stimulus¹

1.1 Introduction

Going back to Hall and Jorgenson (1967), public and macroeconomists have asked how taxes affect investment. The answer is central to the design of countercyclical fiscal policy, since policymakers often use tax-based investment incentives to spur growth in times of economic weakness. Such policies often coincide with disruptions in capital markets, so it is natural to ask how taxes affect investment in the presence of financial frictions. However, the standard theoretical and empirical treatments assume perfect capital markets.² This paper uses recent episodes of investment stimulus to study whether the effect of taxes on

¹Co-authored with James Mahon

²Key theoretical studies include Hall and Jorgenson (1967), Tobin (1969), Hayashi (1982), Abel and Eberly (1994), and Caballero and Engel (1999). Abel (1990) presents a unifying synthesis of the early theoretical literature. Key empirical work includes Summers (1981), Auerbach and Hassett (1992), Cummins, Hassett and Hubbard (1994), Goolsbee (1998), Chirinko, Fazzari and Meyer (1999), Desai and Goolsbee (2004), Cooper and Haltiwanger (2006), and House and Shapiro (2008). Edgerton (2010) relaxes the frictionless assumption but, in contrast to our study, finds mixed results.

investment accords with the standard, frictionless model. We find that, by ignoring financial frictions, the standard analysis overlooks a crucial driver of firm responses to tax policy.

The policy we study, “bonus” depreciation, accelerates the schedule for when firms can deduct from taxable income the cost of investment purchases. Bonus alters the timing of deductions but not their amount, so the economic incentive created by bonus works because future deductions are worth less than current deductions. That is, bonus works because of discounting: firms judge the benefits of bonus by the present discounted value of deductions over time.³ Speeding up the timing of deductions reduces short term taxes, but at the expense of higher taxes in the future. With a reasonable risk-adjusted discount rate, bonus depreciation generates a modest subsidy, so the frictionless model predicts a small effect of bonus on investment.⁴ But in the presence of financial frictions, firms sharply discount future deductions. Thus financial frictions make bonus more appealing, since the difference in today’s tax benefits dwarfs the present value comparison that matters in theory.

We study two episodes of bonus depreciation using a standard difference-in-differences methodology to estimate the effect of these policies. We present three empirical findings. First, bonus depreciation has a substantial effect on investment, much larger than past estimates and much stronger than the conventional wisdom predicts. Estimates of how tax changes affect investment vary, but the consensus prediction is that bonus depreciation has a small positive effect.⁵ In contrast, we find that bonus depreciation raised eligible investment by 17.3 percent on average between 2001 and 2004 and 29.5 percent between 2008 and 2010.

³Summers (1987) states this most clearly: “It is *only* because of discounting that depreciation schedules affect investment decisions. . . .”

⁴Consider a firm making a \$100 investment in computer equipment, which the IRS assigns a depreciation life of five years. Under the normal schedule, a firm can only deduct from taxable income \$20 in the year of the purchase and must spread the remaining \$80 over the next five years. Assuming a seven percent risk-adjusted discount rate, the difference in present values for even full expensing—i.e., 100 percent bonus, which lets the firm deduct \$100 in the purchase year—implies a subsidy of less than four percent.

⁵Cummins, Hassett and Hubbard (1994) study many corporate tax reforms and public company investment data and conclude that tax policy has a strong effect on investment. Using similar data and a different empirical methodology, Chirinko, Fazzari and Meyer (1999) argue that tax policy has a small effect on investment and that Cummins, Hassett and Hubbard (1994) misinterpret their results. Hassett and Hubbard (2002) survey empirical work and conclude that the range of estimates for the user cost elasticity has narrowed to between -0.5 and -1. Surveying this and more recent work, Bond and Van Reenen (2007) decide “it is perhaps a little too early to agree with Hassett and Hubbard (2002) that there is a new ‘consensus’ on the size and robustness of this effect.”

We estimate a user cost elasticity of approximately 1.6, outside the range of estimates of 0.5 to 1 surveyed by Hassett and Hubbard (2002) and more than double the consensus point estimate.⁶

The first part of the paper details this finding and a litany of robustness tests. The research design compares firms at the same point in time whose benefits from bonus differ. Our strategy exploits technological differences between firms in narrowly defined industries. Firms in industries with most of their investment in short duration categories act as the “control group” because bonus only modestly alters their depreciation schedule. This natural experiment separates the effect of bonus from other economic shocks happening at the same time. If the parallel trends assumption holds—if investment growth for short and long duration industries would have been similar absent the policy—then the experimental design is valid.

The key threat to this design is that time-varying industry shocks may coincide with bonus. This risk is limited for four reasons. First, graphical inspection of parallel trends indicates smooth pretrends and a clear, steady break for short and long duration firms during both the 2001 to 2004 and 2008 to 2010 bonus periods. The effects are the same size in both periods, though different industries suffered in each recession. Second, the estimates are stable across many specifications and after including firm-level cash flow controls, industry Q, and flexible industry trends. Controlling for industry-level co-movement with the macroeconomy actually increases our estimates. Third, the estimates pass a placebo test: the effect of bonus on ineligible investment is indistinguishable from zero. Last, for firms making eligible investments, bonus take-up rates (i.e., do firms fill in the bonus box on the tax form?) are indeed higher in long duration industries. For these reasons, spurious factors are unlikely to explain the large effect of bonus.

Firms respond to bonus depreciation as if they apply implausibly high discount rates to investment decisions. This finding is inconsistent with a frictionless model of firm behavior.

⁶In Section 1.4, we collect estimates from past studies of tax reforms. The average user cost elasticity across these studies is 0.69.

In the second part of the paper, we explore alternative models that generate high effective discount rates by adding financial frictions.⁷ One alternative is costly external finance, which raises the total discount rate firms apply to evaluate projects. Another alternative is managerial myopia, which raises effective discount rates by sharply discounting the future relative to the present. Both models prove useful in explaining our findings.

Our second empirical finding is that, consistent with the costly external finance story, financial constraints amplify the effects of investment stimulus. Nearly all prior empirical tests of financial constraints use public firm data, which is problematic because public firms have the best collateral, the strongest banking relationships and broad access to equity and bond markets.⁸ In contrast, we work with an analysis sample of more than 120,000 public and private companies drawn from two million corporate tax returns. Half the firms in our sample are smaller than the smallest firms in Compustat.⁹ Our baseline estimate therefore averages over substantial heterogeneity in firm type, including many firms likely to face financial constraints.

The largest firms in our sample, those most like the firms in past studies, yield estimates in line with the Hassett and Hubbard (2002) range. In contrast, small and medium-sized firms, previously unstudied, show much stronger responses. Building on the differential response by firm size, we perform a split sample analysis using several markers of ex ante financial constraints. In addition to small firms, non-dividend payers and firms with low cash holdings are 1.5 to 2.6 times more responsive than their unconstrained counterparts. Moreover, we find that firms respond by borrowing and cutting dividends. These facts do not match the frictionless model of investment behavior, in which firms divided by financial

⁷We use the term *financial frictions* as an umbrella term over a class of models that generate high effective discount rates. Some of these—such as managerial myopia and agency theory—are not about external finance per se, but refer instead to organizational frictions. These theories have not crossed from the finance literature into standard public and macroeconomics treatments.

⁸Kaplan and Zingales (1997) find that very few of Fazzari, Hubbard and Petersen's (1988a) most constrained firms appear constrained by other measures.

⁹When aggregated, these small firms account for a large amount of economic activity. According to Census tabulations in 2007 (<http://www.census.gov/econ/susb/data/susb2007.html>), firms with less than \$100 million in receipts (around the 80th percentile in our data) account for more than half of total employment and one third of total receipts.

constraint markers do not respond differently to bonus.

Firms with tax losses must wait to realize the benefits of tax breaks. Because many firms in our sample are in a tax loss position when a policy shock occurs, we can ask whether firms value *future* cash windfalls, namely, the larger deductions bonus depreciation provides them in later years. Our third empirical finding is that, consistent with the managerial myopia story, firms *only* respond to investment incentives when the policy immediately generates cash flows. This finding holds even though firms can carry forward unused deductions to offset future taxes, and it cannot be explained by differences in growth opportunities. Furthermore, this fact contradicts a simple model of costly external finance, because firms neglect how the policy affects borrowing in the future.

To confirm the myopia story, we study a second component of the depreciation schedule. Firms making small investment outlays face a permanent kink in the tax schedule, which creates a discontinuous change in marginal investment incentives. This sharp change in incentives induces substantial investment bunching, with many firms electing amounts within just a few hundred dollars of the kink. And when legislation raises the kink, the bunching pattern follows. Consistent with myopia, bunching strongly depends on a firm's current tax status: firms just in positive tax position are far more likely to bunch than firms on the other side of the discontinuity. For a different group of firms and a different depreciation policy, we again find that firms ignore future tax benefits.

These facts do not match the predictions of a frictionless model, which cannot account for the large baseline response, the differential response for constrained firms or the nonresponse for nontaxable firms. The facts point instead toward models in which costly finance matters and current benefits outweigh future benefits. We use an investment model to clarify these findings. The model incorporates costly external finance and managerial myopia into a general model in which the frictionless model of Hayashi (1982) is a special case. These alternative theories make predictions about the discount rate firms apply to future cash flows. The model shows how to combine reduced form estimates to distinguish the frictionless benchmark from costly external finance and managerial myopia.

The general model yields a set of theoretical moments—one comparing constrained and unconstrained firms and one comparing taxable and nontaxable firms—which we can combine with our empirical findings to measure financial frictions. With these comparisons we can estimate the shadow cost of external funds and an implied present versus future discount factor. We estimate the shadow cost of external funds to be between \$0.63 and \$1.61 per dollar and an implied discount factor of 0.84. Combining these results, financially constrained firms act as if \$1 next year is worth just 38 cents today, yielding a total discount rate of 97 percent. Thus accounting for the effect of bonus depreciation on investment requires a major role for financial frictions.

Our paper sits at the intersection of several strands in the economics and finance literatures. Most directly, the paper relates to studies of the effect of taxes on business investment. Our data improve on past studies by including two periods of bonus depreciation; a granular breakdown of eligible investment; a large sample of small, private firms; and better tax variables. Earlier studies pool the effects of different tax reforms, which include depreciation changes, tax rate changes and rule changes regarding corporate form. We focus on one specific policy, bonus depreciation, and carefully dissect how firms respond. House and Shapiro (2008) study the first episode of bonus depreciation using aggregate investment data. Their design compares residuals from a prediction model for investment in short duration (e.g., computers) and long duration (e.g., blast furnaces) categories. We compare firm-level investment growth rates, so that our first difference does not depend on an estimated model. In the literature on salience and taxation, our study offers an example of a strong tax policy effect on economic behavior.¹⁰

The paper also relates to the literature on financial constraints.¹¹ We use depreciation

¹⁰Our evidence is consistent with the strong behavioral response to and salience of the Earned Income Tax Credit (Chetty, Friedman and Saez, 2013). It stands, for instance, in contrast to evidence that individuals react incompletely to obscure taxes (Chetty, Looney and Kroft, 2009) and that business investment does not react to changes in the dividend tax (Yagan, 2013).

¹¹Fazzari, Hubbard and Petersen (1988a) argue that, if firms more likely to be financially constrained respond more strongly to cash flow shocks, then financial constraints are responsible. Subsequent studies make this argument while identifying quasi-experimental variation in cash flows or credit supply (Lamont, 1997; Rauh, 2006; Chaney, Sraer and Thesmar, 2012). We apply this insight to the case of bonus depreciation, which creates

changes as a plausibly exogenous financial constraint shock.¹² Unlike past studies, our instrument also changes the relative price of investment. We use this feature and an investment model to estimate the shadow price of internal funds from the difference between constrained and unconstrained firm elasticities. Our findings imply that incorporating financial frictions adds much explanatory power to neoclassical investment theory.¹³

Cummins, Hassett and Hubbard (1995) and Edgerton (2010) note that tax losses will reduce the incentive of firms to respond to tax changes. The former study uses a sample of sixty loss firms to conclude that losses reduce the effect of tax breaks on investment. The latter maps financial accounting data to a tax account and finds mixed evidence that losses matter.¹⁴ With our data, we can precisely measure whether a firm's current tax position means that the next dollar of investment affects this year's tax bill. Our sample of loss firms includes almost two hundred thousand loss year observations.

The outline of the paper is as follows. Section 1.2 formalizes intuition about how bonus works and develops a set of testable hypotheses, which guide the empirical analysis. Section 1.3 describes the corporate tax data, variable construction and sample selection process. Section 1.4 describes the main empirical strategy for studying bonus depreciation, the identification assumptions and presents results. Section 1.5 uses split sample tests by markers of financial constraints and by tax position to show financial frictions can account for the large baseline effect of bonus. We then develop a set of theoretical moments and combine the empirical results to estimate implied discount rates. Section 1.6 studies substitution margins and external finance responses to bonus. There, we use our heterogeneous size

a shock to the amount of cash firms need to perform their desired investment.

¹²See the conclusion of Fazzari, Hubbard and Petersen (1988a) and Fazzari, Hubbard and Petersen (1988b) for a discussion of how taxes might affect investment in the presence of financial constraints. They focus on average tax rates more generally and do not perform an empirical analysis along these lines.

¹³Hayashi (1982) derives the workhorse neoclassical model. Stein (2003) surveys models in which financial frictions influence investment decisions.

¹⁴Recent work documents large differences between "book" and tax accounts, which introduces the risk of measurement error into such a mapping (see, e.g., Mills, Newberry and Trautman (2002). Edgerton (2010) is very careful with this procedure, but acknowledges that "[one] cannot rule out, however, the possibility that difficulties in measuring firms' taxable status drive the relative unimportance of taxable status observed in the Compustat data."

results to estimate the aggregate investment effect. Section 1.7 discusses policy implications and avenues for future research.

1.2 Hypothesis Development

To direct our empirical analysis, we develop a simple model of investment in the presence of depreciation incentives, financial constraints and heterogeneous tax positions. We modify the neoclassical investment model with adjustment costs (Abel, 1982; Hayashi, 1982) by introducing an external finance wedge and managerial myopia. The model yields a formal hypothesis corresponding to our baseline investment design: investment increases due to bonus depreciation and increases more for industries doing longer lived investment. We develop two further hypotheses corresponding to split sample tests. First, investment responds more when firms face an external finance wedge. Second, investment responds less when firms are in tax loss positions. Here, we focus on the intuition of the model and the mapping from theory to empirical objects and tests. We use a simple one shot static investment model with a reduced form credit wedge.¹⁵ Appendix A.1 derives the hypotheses in an infinite horizon setting with adjustment costs and a dynamic leverage constraint.

Consider a firm making a one shot investment decision. The firm begins with initial profits π_0 and chooses a level of investment I to determine the capital stock and hence future profits.¹⁶ Future profits are given by $\pi(I)$, taxed at the proportional corporate tax rate τ . The firm discounts future flows at risk-adjusted rate r .

The tax code permits the firm to write off the cost of investment over time. The value of these deductions depends on the tax rate and how the schedule interacts with the firm's

¹⁵This wedge is a reduced form model of a set of capital market frictions, which might reflect, e.g., costly monitoring problems or adverse selection (Stein, 2003).

¹⁶Normalize the price of investment to one.

discount rate. We collapse the stream of future depreciation deductions owed for investment:

$$z^0(\beta) = D_0 + \beta \sum_{t=1}^T \frac{1}{(1+r)^t} D_t, \quad (1.1)$$

where D_t is the allowable deduction per dollar of investment in period t and T is the class life of investment. $z^0(\beta)$ measures the present discounted value of one dollar of investment deductions before tax. If the firm can immediately deduct the full dollar, then z^0 equals one. Because of discounting, z^0 is lower for longer lived items (i.e., items with greater T), which forms the core of our identification strategy.

In general, the stream of future deductions depends on future tax rates and discount rates. Our empirical analysis assumes the effective tax rate does not change over time, except when the firm is nontaxable.¹⁷ For discount rates, we begin by assuming a risk-adjusted rate of seven percent to compute z^0 in the data, which enables comparison to past work. We then relax this assumption in Section 1.5.3 when we estimate an implied discount rate. β is an additional discount term between zero and one, which reflects the possibility of myopia.¹⁸ We use our heterogeneity analysis to identify this term separately.¹⁹

Bonus depreciation, the policy we study in our empirical analysis, allows the firm to deduct a per dollar bonus, θ , at the time of the investment and then depreciate the remaining $1 - \theta$ according to the normal schedule:

$$z(\beta) = \theta + (1 - \theta)z^0(\beta) \quad (1.2)$$

At different points in time, Congress has set θ equal to 0, 0.3, 0.5 or 1. We use these policy

¹⁷We use the top statutory tax rate in the set of specifications requiring a tax rate. This is an upper bound on the more realistic effective marginal tax rate, which in turn depends on tax rate progressivity and the level of other expenses relative to taxable income. See, e.g., Graham (1996, 2000) for a method tracing out the marginal tax benefit curve. The policies we study will increase the use of investment as a tax shield regardless of where the firm is on this marginal benefit curve. Except when current and all future taxes are zero, bonus increases the marginal tax benefit of investment.

¹⁸The myopia model is closer to Akerlof (1991) and Laibson (1997) than it is to the model of managerial myopia in Stein (1989). Stein's (1989) model of managerial myopia specifically refers to the incentive to boost current earnings as a way of signaling high quality to the stock market. We use the term to reflect any motive to boost current earnings and neglect projects with long term payoffs and short term costs.

¹⁹We assume that β explicitly applies to depreciation deductions, which have both a present and future component. In the case of profits, which only arrive in the future, we assume π incorporates β implicitly.

shocks to identify the effect of bonus depreciation on investment. Industries differ by average z^0 prior to bonus, providing the basis for identification in a difference-in-differences setup with continuous treatment.

We further generalize z by incorporating a nontaxable state. When the next dollar of investment does not affect this year's tax bill, then the firm must carry forward the deductions to future years.²⁰ Our general z reflects this case:

$$z(\beta, \gamma) = \gamma z(\beta) + (1 - \gamma)\beta\phi z(1), \quad (1.3)$$

where $\gamma \in \{0, 1\}$ is an indicator for current tax state and ϕ is a discount factor that reflects both the expected arrival time of the taxable state and the discount rate applied to the future and subsequent periods when the firm switches. Note that for the nontaxable firm, β applies to all future deductions. Even when β equals one, ϕ is less than one, so the value of these deductions are lower when the firm is nontaxable. We measure ϕ in the data and apply our split sample results to determine whether we can justify these findings in a model without myopia.

External finance matters for all investment exceeding current cash flow. During the investment period, the firm faces an external finance wedge that is linear in expenses net of cash flows, that is,

$$c(I) = \lambda [(1 - \tau z)I - (1 - \tau)\pi_0], \quad (1.4)$$

where λ can be thought of as the shadow price on a borrowing constraint that may or may not bind now or in the future. Thus, a dollar of cash inside the firm is worth $1 + \lambda$.²¹ We include z in the net expense term and not just the first year deduction, to capture the influence of depreciation deductions on future taxes and thus future borrowing. While

²⁰This assumes that “carrybacks”—in which firms apply unused deductions this year against past tax bills—have been exhausted or ignored. In ongoing work, we find that carryback take-up rates are surprisingly low.

²¹Note that because we have assumed a linear external finance function, there will be no direct effect of cash flows on investment, that is, the investment-cash flow sensitivity is zero. Because each dollar of investment can only generate at most 35 cents of cash back, these policies cannot operate mainly through a direct cash windfall channel.

bonus depreciation relaxes the current constraint through reducing this year's tax bill, it does so at the expense of higher future taxes. The net effect is to reduce the present discounted borrowing costs for the firm. However, if myopia plays a role (that is, for low β), then only the current year change will matter. The two models thus yield different predictions for constrained, nontaxable firms: constrained, myopic firms respond much less to bonus when nontaxable than do constrained, farsighted firms. This is the feature we use to distinguish costly external finance from myopia models, which are otherwise observationally equivalent.

We derive a condition for optimal investment. Though the problem occurs over time, we can write it as a static one shot investment problem by discounting future flows to the present. Discarding elements not involving investment, the firm's objective is

$$\max_I \left\{ \frac{(1-\tau)\pi(I)}{1+r} - (1-\tau z)I - \lambda(1-\tau z)I \right\} \quad (1.5)$$

Here, we assume π is weakly concave, which ensures that the problem yields a unique interior solution.

The first order condition for optimal investment is

$$(1-\tau)\pi'(I^*) = (1+r)(1+\lambda)(1-\tau z). \quad (1.6)$$

Intuitively, the investment decision trades off the after-tax future benefits of the marginal dollar of investment against its price (normalized to one) and the marginal external finance cost, less the marginal benefit due to depreciation deductions. Deductions lower the hurdle rate for investment both through their net present value and through relaxing the external finance constraint. With costly external finance, optimal investment is strictly lower than in the frictionless case or when inside cash can cover all investment expenses (i.e., when $\lambda = 0$).

We derive three testable hypotheses from the model. The first concerns the average effect of bonus depreciation on investment, while the latter two concern heterogeneous effects by the presence of costly external finance and by tax position. Bonus depreciation increases

the present value of deductions, reducing the price of investment. Thus bonus depreciation should increase investment. Each hypothesis builds on the comparative static with respect to the bonus parameter θ . In the Appendix, we show that investment is increasing in θ .

Hypothesis 1 *Investment responds more strongly to bonus depreciation for industries with more investment in longer lived eligible items. That is, $\partial^2 I / \partial \theta \partial z^0 < 0$.*

Bonus depreciation works through increasing θ . Hypothesis one concerns the basic effect of this policy on investment. The more delayed the normal depreciation schedule is, the more generous bonus will be. Longer lived items like telephone lines and heavy manufacturing equipment have a more delayed baseline schedule than short lived items like computers (i.e., $z_{\text{Long}}^0 < z_{\text{Short}}^0$). Thus, industries that buy more long lived equipment see a larger relative price cut when bonus happens.

Our second hypothesis concerns how the investment response varies with costly external finance.

Hypothesis 2 *Investment responds more strongly to bonus depreciation for financially constrained firms. That is, $\partial^2 I / \partial \theta \partial \lambda > 0$.*

For financially constrained firms, bonus depreciation both reduces the price of investment and reduces how much they have to borrow. The effective price change is thus larger for constrained firms. We use several proxies for ex ante financial constraints—firm size, dividend payment activity and liquid asset positions—to test for a difference in elasticities between constrained and unconstrained firms. If financial constraints are unimportant, then we should not find a consistent, systematic difference in elasticities for groups of firms based on these proxies. We can use the difference in coefficients between constrained and unconstrained firms to ask what the implied external finance spread. We formalize and implement this intuition in Section 1.5.3.

Our third hypothesis concerns how the investment response varies with the firm's current tax position.

Hypothesis 3 *Investment responds more strongly to bonus depreciation for firms with current-year taxable income. That is, $\partial I / \partial \theta|_{\gamma=1} > \partial I / \partial \theta|_{\gamma=0}$.*

Hypothesis three emerges in any model with some positive discounting, since future benefits are worth less than immediate benefits. The main value of the comparison between taxable and nontaxable groups derives from the calibration it offers. We can calibrate the expected arrival of the taxable state for nontaxable firms and ask whether the difference between elasticities for taxable and nontaxable firms requires some myopia (i.e., $\beta < 1$).

1.3 Business Tax Data

The analysis in this paper uses the most complete dataset yet applied to study business investment incentives.²² The data include detailed information on equipment and structures investment, offering a finer breakdown than previously available for a broad class of industries. The sample includes many small, private firms and all of the largest US firms, which enables the heterogeneity analysis we use to document financial constraints. Because the data come from corporate tax returns, we can separate firms based on whether the next dollar of investment affects this year's taxes. This allows a split sample analysis that can distinguish the myopia model from a simple model of costly external finance. In this section, we describe where these data come from and the analysis sample, as well as how we map the theory into empirical objects.

Sampling Process. Each year, the Statistics of Income (SOI) division of the IRS Research, Analysis and Statistics unit produces a stratified sample of approximately 100,000 unaudited corporate tax returns.²³ Stratification occurs by form type,²⁴ total assets, and proceeds.

²²Yagan (2013) uses these data to study the 2003 dividend tax cut. Kitchen and Knittel (2011) use these data to describe general patterns in bonus and Section 179 take-up.

²³Details come from <http://www.irs.gov/pub/irs-soi/08cosec3ccr.pdf>.

²⁴For example, C corporations file form 1120 and S corporations file form 1120S. Other form types include real estate investment trusts, regulated investment companies, foreign corporations, life insurance companies, and property and casualty insurance companies. Sampling frequencies reflect the distribution of these types in

Each sample year includes returns with accounting periods ending between July of that year and the following June. When necessary, we recode the tax year to align with the implementation of the policies studied in this paper. In 2008, the sample represented about 1.8 percent of the total population of 6.4 million C and S corporation returns.

SOI uses these samples to generate annual publications documenting income characteristics. The BEA uses them to finalize national income statistics. In addition, the Treasury's Office of Tax Analysis (OTA) uses the sample to perform policy analysis and revenue estimation. To enable these aggregate statistics, SOI carefully reports sampling weights which reflect each observation's sampling frequency. Our aggregate estimates incorporate these weights. Any corporation selected into the sample in a given year will be selected again the next year, providing it continues to fall in a stratum with the same or higher sampling rate. Shrinking firms are resampled at a lower rate, which introduces sampling attrition. We address this attrition in several ways, including a nonparametric reweighting procedure for figures and through assessing the robustness of our results in a balanced panel.

Analysis Samples, Variable Definitions and Summary Statistics. We create a panel by linking the cross sectional SOI study files using firm identifiers.²⁵ The raw dataset has 1.84 million rows covering the years from 1993 to 2010. There are 355 thousand distinct firms in this dataset, 19,711 firms with returns in each year of the sample and 62,478 firms with at least 10 years of returns. Beginning with the sample of firms with valid data for each of the main data items analyzed, we keep firm-years satisfying the following criteria: (a) having non-zero total deductions or non-zero total income²⁶ and (b) having an attached investment

the population. We focus on 1120 and 1120S, which cover the bulk of business activity in industries making equipment investments.

²⁵We thank OTA staff for providing the data crosswalk.

²⁶Knittel et al. (2011) use a similar "de minimus" test to select business entities that engage in "substantial" business activity.

form.²⁷ In addition, we exclude partial year returns, which occur when a firm closes or changes its fiscal year. To analyze bonus depreciation, we exclude firms potentially affected by Section 179, a small firm investment incentive which we analyze separately. Our main bonus analysis sample consists of all firms with average eligible investment greater than \$100,000 during years of positive investment.²⁸ This sample consists of 820,769 observations for 128,151 distinct firms.

This section describes the economic concepts underlying the variables we study. **Eligible investment**, our main variable of interest, includes expenditures for all equipment investment put in place during the current year for which bonus and Section 179 incentives apply.²⁹ We conduct separate analyses for intensive and extensive margin responses. The intensive margin variable is the logarithm of eligible investment. The extensive margin variable is an indicator for positive eligible investment. We aggregate this indicator at the industry level and transform it into a log odds ratio³⁰ for our empirical analyses. In some specifications, we use an alternative measure of investment, which is eligible investment divided by lagged **capital stock**. Capital stock is the reported book value of all tangible, depreciable assets. **Sales** equals operating revenue and **assets** equals total book assets. **Total debt** equals the sum of non-equity liabilities excluding trade credit. **Liquid assets** equals cash and other liquid securities. **Payroll** equals non-officer wage compensation. **Rents** equals lease and rental expenses. **Interest** equals interest payments.

Our main policy variable of interest, $z_{N,t}$, is the present discounted value of one dollar of deductions for eligible investment. In each non-bonus year, we compute the share of

²⁷Form 4562 is the tax form that corporations attach to their return to claim depreciation deductions on new and past investments. An entity that claims no depreciation deductions need not attach form 4562. It is likely that these firms do not engage in investment activity, and so their exclusion should not affect the interpretation of results.

²⁸The relevant threshold for Section 179 was \$25,000 until 2003, when it increased to \$100,000. In 2008, it increased to \$250,000 and then to \$500,000 in 2010. Using alternative thresholds in the range from \$50,000 to \$500,000 does not alter the results.

²⁹Section 179 and bonus rules differ slightly, in that Section 179 also applies to used equipment purchases, while bonus only applies to new equipment. The form does not require firms to list used purchases separately.

³⁰I.e., we use $\log(\frac{p}{1-p})$ as our measure of the extensive margin.

eligible investment a firm reports in each category.³¹ We use these shares and the present value of one dollar of eligible investment for each category to construct a weighted average, firm-level z . Category z 's come from applying a seven percent discount rate to the pertinent deduction schedule, while assuming a six-month convention for the purchase year.^{32,33} We follow compute z_N at the four-digit NAICS industry level as the simple average of the firm-level z 's across non-bonus years prior to 2001.³⁴ In bonus years, we adjust z by the size of the bonus. If θ is the additional expense allowed per dollar of investment (e.g., $\theta = .3$ for 2001), then $z_{N,t|\theta_t} = \theta_t + (1 - \theta_t) \times z_N$. The interaction between the time series variation in θ and the cross sectional variation in z_N delivers the identifying variation we use to test our three hypotheses.

Table 1.1 collects summary statistics for the sample in our bonus depreciation analysis. The average observation has \$6.8 million in eligible investment, \$180 million in sales and \$27 million in payroll. The size distribution of corporations is skewed, with median eligible investment of just \$370 thousand and median revenues of \$26 million. The average net present value of depreciation allowances, $z_{N,t}$, is 0.88 in non-bonus years, implying that eligible investment deductions for a dollar of investment are worth eighty-eight cents to the average firm. $z_{N,t}$ increases to an average of 0.94 during bonus years. Cross sectional differences in $z_{N,t}$ are similar in magnitude to the change induced by bonus, with $z_{N,t}$ varying from 0.87 at the tenth percentile to 0.94 at the ninetieth. The first year deduction, $\theta_{N,t}$, increases from an average of 0.18 in non-bonus years to 0.58 in bonus years.

³¹Specifically, 3-, 5-, 7-, 10-, 15-, and 20-year Modified Accelerated Cost Recovery System (MACRS) property and listed property.

³²The category deduction schedules are available in IRS publication 946. We use a seven percent rate as a frictionless benchmark that is likely larger than the rate firms should be using, which will tend to bias our results downward. Summers (1987) argues that firms should apply a discount rate close to the risk-free rate for depreciation deductions. Seven percent is the largest discount rate House and Shapiro (2008) apply when computing the value of bonus depreciation.

³³The six-month convention is applied because on average the property is in place for only half of the first year.

³⁴Like Cummins, Hassett and Hubbard (1994) and Edgerton (2010), we proxy for the firm-level benefit of bonus depreciation with an industry measure of policy benefits. Unlike these studies, our measure derives directly from tax data, reducing measurement error. It is possible to apply the same strategy at the firm level. This approach does not alter our findings.

The difference in z 's over time of just six cents per dollar before tax translates into a benefit of just over two cents after tax, which is why some authors claim the effect of bonus on investment should be small. However, if the discount rate firms apply to future deductions includes a large external finance wedge or myopia, then this two cent difference can increase to as much as the forty cent difference in average θ 's.

Table 1.1: *Statistics: Bonus Analyses*

	Mean	P10	Median	P90	Count
Investment Variables					
Investment (000s)	6,786.87	0.81	367.59	5,900.17	818,576
log(Investment)	6.27	4.10	6.14	8.81	735,341
Investment/Lagged Capital Stock	0.10	0.00	0.05	0.27	637,243
$\Delta \log(\text{Capital Stock})$	0.08	-0.05	0.05	0.33	637,278
log(Odds Ratio _N)	1.28	0.54	1.34	2.05	818,107
Other Outcome Variables					
$\Delta \log(\text{Debt})$	0.04	-0.37	0.03	0.56	642,546
$\Delta \log(\text{Rent})$	0.08	-0.38	0.04	0.66	574,305
$\Delta \log(\text{Wage Compensation})$	0.06	-0.21	0.05	0.40	624,918
log(Structures Investment)	5.02	2.13	4.98	8.10	389,232
Policy Variables					
$z_{N,t}$	0.90	0.87	0.89	0.94	818,576
Characteristics					
Assets (000s)	403,597.2	3,267.96	24,274.82	327,301.6	818,576
Sales (000s)	180,423.8	834.65	25,920.92	234,076.1	818,576
Capital Stock (000s)	89,977.09	932.00	7,214.53	80,122.69	818,576
Net Income Before					
Depreciation (000s)	15,392.59	-2,397.92	1,474.65	17,174.55	818,576
Profit Margin	0.17	-0.07	0.05	0.68	777,968
Wage Compensation (000s)	26,826.36	372.09	4,199.88	38,526.46	818,576
Cash Flow/Lagged Capital Stock	0.05	-0.09	0.03	0.26	647,617

Notes: This table presents summary statistics for analysis of bonus depreciation. To preserve taxpayer anonymity, "percentiles" are presented as means of all observations in the $(P - 1, P + 1)$ th percentiles. Investment is bonus eligible equipment investment. $z_{N,t}$ is the weighted present value for a dollar of eligible investment expense at the four-digit NAICS level, with weights computed using shares of investment in each eligible category. The odds ratio is defined at the four-digit NAICS level as the fraction of firms with positive investment divided by the fraction with zero investment. Cash flow is net income before depreciation after taxes paid. Ratios are censored at the one percent level.

It is helpful to give a sense of the groups being compared, because our identification will be based on assuming that industry-by-year shocks are not confounding the trends between industry groups. The five most common three-digit industries (NAICS code) in the bottom

three z_N deciles are: motor vehicle and parts dealers (441), food manufacturing (311), real estate (531), telecommunications (517), and fabricated metal product manufacturing (332). In the top three deciles are: professional, scientific and technical services (541), specialty trade contractors (238), computer and electronic product manufacturing (334), durable goods wholesalers (423), and construction of buildings (236). Neither group of industries appears to be skewed toward a spurious relative boom in the low z group. The telecommunications industry suffered unusually during the early bonus period as did real estate in the later period. Both industries are in the group for which we observe a larger investment response due to bonus.

1.4 The Effect of Bonus Depreciation on Investment

We begin with a test of Hypothesis 1, which predicts that investment responds more strongly to bonus depreciation for industries with more investment in longer lived eligible items. In both bonus periods we study, we estimate large responses to bonus depreciation. The estimates are similar in both periods. We assess the key risk of this design—that time-varying industry shocks confound our estimates—using a variety of specifications, a placebo test and differences in policy salience across space.

Policy Background. House and Shapiro (2008) provide a detailed discussion of the baseline depreciation schedule³⁵ and legislative history of the first round of bonus depreciation. Kitchen and Knittel (2011) provide a brief legislative history of the second round.³⁶ Appendix A.2 summarizes the relevant legislation.

In 2001, firms buying qualified investments³⁷ were allowed to immediately write off 30 percent of the cost of these investments. The bonus increased to 50 percent in 2003

³⁵Known as the Modified Accelerated Cost Recovery System, or MACRS.

³⁶See also the Treasury’s “Report to The Congress on Depreciation Recovery Periods and Methods” (2000).

³⁷Depreciable tangible personal property with class life of twenty years or less, purchased for use in the active conduct of a trade or business. Used equipment was excluded.

and expired at the end of 2004. In 2008, 50 percent bonus depreciation was reinstated. It was later extended to 100 percent bonus for tax years ending between September 2010 and December 2011.³⁸ The policies applied to equipment and excluded most structures.³⁹

The policies were intended as economic stimulus. In the words of Congress, “increasing and extending the additional first-year depreciation will accelerate purchases of equipment, promote capital investment, modernization, and growth, and will help to spur an economic recovery” (Committee on Ways & Means, 2003, p. 23). To avoid encouraging firms to delay investment until the policy came online, legislators announced that the policy would apply retroactively to include the time when the policy was under debate. Although the first bonus legislation passed in early 2002, firms anticipating policy passage would have begun responding in the fourth quarter of 2001. We therefore include firm-years with the tax year ending within the legislated window in our treatment window.

Whether firms perceived the policy as temporary or permanent is a subject of debate. The initial bill branded the policy as temporary stimulus, slating it to expire at the end of 2004, which it did. For this reason, House and Shapiro (2008) assume firms treat the policy as temporary. In contrast, Desai and Goolsbee (2004) cite survey evidence indicating that many firms expected the provisions to continue, and our empirical analysis in Section 1.6 offers little evidence of intertemporal shifting.⁴⁰ Expecting the policy to be temporary is critical to House and Shapiro (2008), because their exercise relies upon how policies approximated as instantaneous interact with the duration of investment goods approximated as infinitely lived. Our design relies much less on this assumption. In our model, costly external finance and myopia amplify the effects of both temporary and permanent policies. And our cross

³⁸In the first bonus period, property had to be put in place after September 10, 2001 and before January 1, 2005. The start date for the second bonus period was December 31, 2007 and the end date was December 31, 2011.

³⁹These provisions coincided with an increase in the Section 179 allowance for small investments from \$24,000 to \$100,000 in 2003, from \$125,000 to \$250,000 in 2008, and from \$250,000 to \$500,000 in 2010. In our main sample, we exclude firms with mean eligible investment greater than \$100,000. Altering this threshold does not change our results.

⁴⁰Note that when production functions exhibit constant returns to scale, the effect of temporary and permanent policies on investment will be the same (Abel, 1982).

sectional identification relies much less on the response of the longest lived investment goods.

Empirical Setup. Bonus depreciation provides a temporary reduction in the price and a temporary increase in the first year deduction for eligible investment goods. Eligible items are classified for deduction profiles over time based on their useful life. Identification builds upon the idea that some industries benefited more from these cuts by virtue of having longer duration investment patterns, that is, by having more investment in longer class life categories. This cross-sectional variation permits a within-year comparison of investment growth for firms in different industries.⁴¹ The policy variation is at the industry-by-year level, so the key identifying assumption is that the policies are independent of other industry-by-year shocks. Several robustness tests validate this assumption.

The regression framework implements the difference-in-differences (DD) specification,

$$f(I_{it}, K_{i,t-1}) = \alpha_i + \beta g(z_{N,t}) + \gamma X_{it} + \delta_t + \varepsilon_{it}, \quad (1.7)$$

where $z_{N,t}$ is measured at the four-digit NAICS industry level and increases temporarily during the bonus years. The specific additive form we adopt in (1.7) for the unobserved firm-level components, α_i , can only be valid for a particular class of investment functions. For example, if valid in levels, the design cannot be valid in logs. The investment data summarized in Table 1.1 is highly skewed with a mean of \$6.8 million and a median of just \$368 thousand. Thus, a multiplicative unobserved effect (that is, $I_i = A_i I^*(z)$) is the most likely empirical model for investment levels. This delivers an additive model in logarithms, which is the approach we pursue below. Because approximately eight percent of our observations for eligible investment are equal to zero, we supplement the intensive margin logs approach with a log odds model for the extensive margin. We measure the log

⁴¹This methodological approach was first applied in Cummins, Hassett and Hubbard (1994). See also Cummins, Hassett and Hubbard (1996), Desai and Goolsbee (2004), House and Shapiro (2008) and Edgerton (2010).

odds ratio as $\log(P[I > 0]/(1 - P[I > 0]))$ at the four-digit industry level.⁴²

Studies often use an alternative empirical specification for $f(I, K)$, where investment is scaled by lagged assets or lagged capital stock. We prefer log investment for four reasons. First, small firms are not always required to disclose balance sheet information, so requiring reported assets would reduce our sample frame. Second, and related to the first reason, requiring two consecutive years of data for a firm-year reduces our sample by fifteen percent. Third, there is some concern that balance sheet data on tax accounts are not reported correctly for consolidated companies due to failure to net out subsidiary elements.⁴³ Measurement error in the scaling variable introduces non-additive measurement error into the dependent variable. Last, with multiple types of capital, the scaling variable might not remove the unobserved firm effect from the model. This is especially a concern because we cannot measure a firm's stock of eligible capital and because firms vary in the share of total investments made in eligible categories.⁴⁴ While we prefer the log investment model, we also report results using investment scaled by lagged capital stock, which allows comparison to past studies.

Graphical Evidence. Figure 1.1 presents a visual implementation of this research design. To allow a comparison that matches a regression analysis with fixed effects and firm-level covariates, we construct residuals from a two-step regression procedure. First, we nonparametrically reweight (i.e., Dinardo, Fortin and Lemieux (1996) reweight) the group-by-year distribution within ten size bins based on assets crossed with ten size bins based on sales.⁴⁵ This procedure addresses sampling frame changes over time, which cause instability

⁴²An alternative specification, with the odds ratio replaced by $P[I > 0]$, works as well. However, the logs odds ratio has better statistical properties (e.g., a more symmetric distribution).

⁴³Mills, Newberry and Trautman (2002) analyze balance sheet accounting in tax data and document difficulties in reconciling these accounts with book accounts.

⁴⁴Abel (1990) notes that this issue and other violations of linear homogeneity can lead to spurious conclusions (e.g., a reversed investment-Q relationship).

⁴⁵The bins are set based on the size distribution in 2000.

in the aggregate distribution.⁴⁶ In the second step, we run cross sectional regressions each year of the outcome variable on an indicator for treatment group—either long duration or short duration—and a rich set of controls, including ten-piece splines in assets, sales, profit margin and age. We plot the residual group means from these regressions.⁴⁷

We compare mean investment in calendar time for the top and bottom three deciles of the investment duration distribution.⁴⁸ Long duration industries show growth well above that of the short duration industries, with this difference only appearing in the bonus years. The difference between the slopes of these two lines in any year gives the difference-in-differences estimate between these groups in that year. The other years provide placebo tests of the natural experiment and indicate no false positives.

Statistical Results and Economic Magnitudes. Table 1.2 presents regressions of the form in (1.7), where $f(I_{it}, K_{i,t-1})$ equals $\log(I_{it})$ in the intensive margin model, $\log(P_N[I_{it} > 0]/(1 - P_N[I_{it} > 0]))$ in the extensive margin model, and $I_{it}/K_{i,t-1}$ in the user cost model; and $g(z_{N,t})$ equals $z_{N,t}$ in the intensive and extensive margin models and $(1 - \tau z_{N,t})/(1 - \tau)$ in the user cost model.⁴⁹ The baseline specification includes year and firm fixed effects. Standard errors are clustered at the firm level in the intensive margin and user cost models.⁵⁰ Because log odds ratios are computed at the industry level, standard errors in the extensive margin model are clustered at the industry level.

The first column reports an intensive margin semi-elasticity of investment with respect

⁴⁶During the period we study, the size of the sample frame changed twice due to budgetary constraints.

⁴⁷To align the first year of each series and ease comparison of trends, we subtract from each dot the group mean in the first year and add back the pooled mean from the first year. All means are count weighted.

⁴⁸Deciles are computed at the industry level.

⁴⁹ τ is set to 35 percent, the top statutory tax rate for all firms.

⁵⁰This is consistent with recent work (e.g., Desai and Goolsbee (2004), Edgerton (2010), Yagan (2013)) and enables us to compare our confidence bands to past estimates. The implicit assumption that errors within industries are independent is strong, for the same reason that Bertrand, Duflo and Mullainathan (2004) criticize papers that cluster at the individual level when studying state policy changes. Our results in this section are robust to industry clustering, as are the tax splits in the next section. In the financial constraint splits regressions, we discuss which inferences are robust to this more conservative structure. We are not aware of other studies that restrict inference in this way and still show that taxes affect investment.

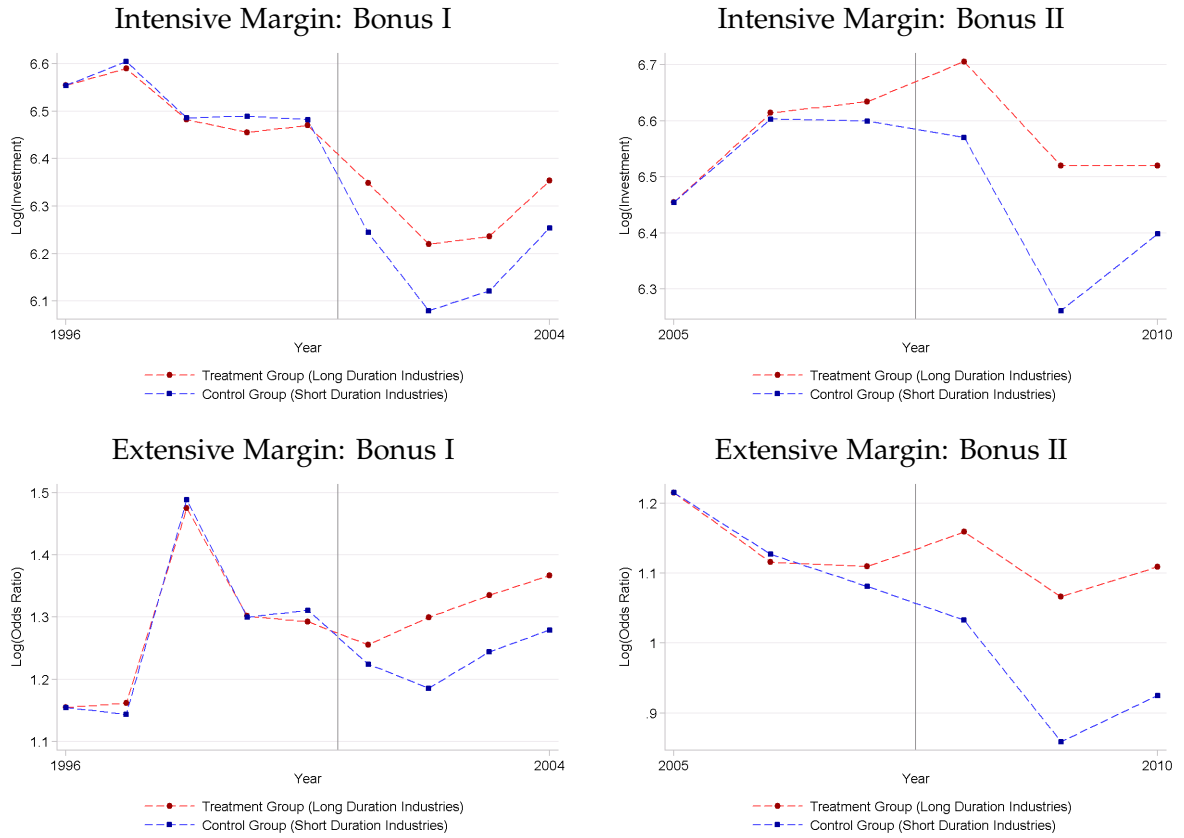


Figure 1.1: Calendar Difference-in-Differences

Notes: Panel (a) plots the average logarithm of eligible investment over time for groups sorted according to their industry-based treatment intensity. Treatment intensity depends on the average duration of investment, with long duration industries (treatment groups) seeing a larger average price cut due to bonus than short duration industries (control groups). Panel (b) plots the industry-level log odds ratio for the probability of positive eligible investment, thus offering a measure of the extensive margin response. The treatment years for Bonus I are 2001 through 2004 and 2008 through 2010 for Bonus II. In these years, the difference between changes in the red and the blue lines provides a difference-in-differences estimator for the effect of bonus in that year for those groups. The earlier years provide placebo tests and a demonstration of parallel trends. The averages plotted here result from a two-step regression procedure. First, we nonparametrically reweight the group-by-year distribution (i.e., Dinardo, Fortin, and Lemieux (1996) reweight) within ten size bins based on assets crossed with ten size bins based on sales to address sampling frame changes over time. Second, we run cross sectional regressions each year of the outcome variable on an indicator for treatment group and a rich set of controls, including ten-piece splines in assets, sales, profit margin and age. We plot the residual group means from these regressions. To align the first year of each series and ease comparison of trends, we subtract from each dot the group mean in the first year and add back the pooled mean from the first year. All means are count weighted.

to z of 3.7, an extensive margin semi-elasticity of 3.8 and a user cost elasticity of -1.6 . The average change in $z_{N,t}$ was 4.7 cents during the early bonus period and 8 cents during the later period, implying average investment increases of $17.3 (= 3.69 \times 4.7)$ and $29.5 (=$

3.69×8) log points, respectively. These predictions should not be confused with the aggregate effect of the policy, because they are based on equal-weighted regressions which include many small firms. They only provide an informative aggregate prediction under the strong assumption that the semi-elasticity is independent of firm size.

In the second column, including a control for contemporaneous cash flow scaled by lagged capital does not alter the estimates. Columns three and four show a similar semi-elasticity for both the early and late episodes. Column five controls for fourth order polynomials in each of assets, sales, profit margin and firm age, as well as industry average Q measured from Compustat at the four-digit level. Column six adds quadratic time trends interacted with two-digit NAICS industry dummies, which causes the estimated semi-elasticity to increase.⁵¹ These alternative control sets do not challenge our main finding: the investment response to bonus depreciation is robust across many specifications.

Appendix Table A.1.2 collects from other studies estimates that we can compare to our user cost model. Like our study, each one uses tax reforms crossed with industry characteristics to estimate the effect of taxes on investment. Panel (a) of Table 1.3 plots the estimates from these studies with confidence bands, highlights the consensus range, and compares them to our estimate. The average user cost elasticity across these studies is 0.69, which falls within Hassett and Hubbard (2002)'s consensus range of 0.5 to 1, but is less than half our estimate of 1.60.⁵²

Additional Robustness and Policy Salience. The calendar time plot in Figure 1.1 provides several visual placebo tests through inspection of the parallel trends assumption in non-bonus years. Because bonus depreciation excludes very long lived items (i.e., structures),

⁵¹We can replace the quadratic time trends with increasingly nonlinear trends or two digit industry-by-time fixed effects. We can also replace the time trends with two-digit industry interacted with log GDP or GDP growth. In each case, the estimates increase. This suggests that omitted industry-level factors bias our estimates downward. Consistent with this story, Dew-Becker (2012) shows that long duration investment falls more during recessions than short duration investment.

⁵²In an investment model, the elasticity of investment with respect to the net of tax rate, $1 - \tau z$, equals the price elasticity and interest rate elasticity, derived in Appendix A.1. Our empirical model delivers an elasticity of 7.2. We are not aware of easily comparable estimates for prices or interest rates.

Table 1.2: Investment Response to Bonus Depreciation

Intensive Margin: LHS Variable is log(Investment)						
	(1)	(2)	(3)	(4)	(5)	(6)
$z_{N,t}$	3.69*** (0.53)	3.78*** (0.57)	3.07*** (0.69)	3.02*** (0.81)	3.73*** (0.70)	4.69*** (0.62)
$CF_{it}/K_{i,t-1}$		0.44*** (0.016)				
Observations	735341	580422	514035	221306	585914	722262
Clusters (Firms)	128001	100883	109678	63699	107985	124962
R ²	0.71	0.74	0.73	0.80	0.72	0.71
Extensive Margin: LHS Variable is log($P(\text{Investment} > 0)$)						
	(1)	(2)	(3)	(4)	(5)	(6)
$z_{N,t}$	3.79** (1.24)	3.87** (1.21)	3.12 (2.00)	3.59** (1.14)	3.99* (1.69)	4.00*** (1.13)
$CF_{it}/K_{i,t-1}$		0.029** (0.0100)				
Observations	803659	641173	556011	247648	643913	803659
Clusters (Industries)	314	314	314	274	277	314
R ²	0.87	0.88	0.88	0.93	0.90	0.90
User Cost: LHS Variable is Investment/Lagged Capital						
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{1-t_c z}{1-t_c}$	-1.60*** (0.096)	-1.53*** (0.095)	-2.00*** (0.16)	-1.42*** (0.13)	-2.27*** (0.14)	-1.50*** (0.10)
$CF_{it}/K_{i,t-1}$		0.043*** (0.0023)				
Observations	637243	633598	426214	211029	510653	631295
Clusters (Firms)	103890	103220	87939	57343	90145	103565
R ²	0.43	0.43	0.48	0.54	0.45	0.44
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	Yes	No
Industry Trends	No	No	No	No	No	Yes

Notes: This table estimates regressions of the form

$$f(I_{it}, K_{i,t-1}) = \alpha_i + \beta g(z_{N,t}) + \gamma X_{it} + \delta_t + \varepsilon_{it}$$

where I_{it} is eligible investment expense and $z_{N,t}$ is the present value of a dollar of eligible investment computed at the four-digit NAICS industry level, taking into account periods of bonus depreciation. Regression (2) augments the baseline specification with current period cash flow scaled by lagged capital. Column (3) focuses on the early bonus period and column (4) focuses on the later period. Column (5) controls for four-digit industry average Q for public companies and quartics in assets, revenues, profit margin and firm age. Column (6) includes quadratic time trends interacted with two-digit NAICS industry dummies. Ratios are censored at the one percent level. Standard errors clustered at the firm level are in parentheses.

Table 1.3: *Investment Response to Bonus Depreciation: Robustness*

	Structures		Net Investment		Has Bonus		Salience Split	
	Basic	Trends	Basic	Trends	Basic	Trends	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$z_{N,t}$	0.52 (0.78)	1.10 (0.98)	1.07*** (0.23)	0.78*** (0.22)	0.95*** (0.13)	1.15*** (0.16)	5.09*** (0.87)	1.56 (0.94)
Observations	389232	381921	637278	631680	818576	804128	211390	215205
Clusters (Firms)	92351	90166	103447	103147	128150	125534	29627	30836
R ²	0.59	0.60	0.27	0.27	0.61	0.61	0.70	0.70
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Trends	No	Yes	No	Yes	No	Yes	No	No

Notes: This table estimates regressions of the form

$$Y_{it} = \alpha_i + \beta z_{N,t} + \delta_t + \varepsilon_{it}$$

where Y_{it} is either the logarithm of structures investment (columns (1) and (2)), log growth in capital stock (columns (3) and (4)), an indicator for take-up of bonus depreciation ((5) and (6)), or the logarithm of eligible investment ((7) and (8)). $z_{N,t}$ is the present value of a dollar of eligible investment computed at the four-digit NAICS industry level, taking into account periods of bonus depreciation. Columns (1), (3), (5), (7) and (8) implement the baseline specification in table 1.2. Columns (2), (4) and (6) include quadratic time trends interacted with two-digit NAICS industry dummies. Columns (7) and (8) split the sample into the top and bottom three deciles according to local geographic salience of the depreciation schedule. We proxy for local salience using frequency of bunching by small firms at the Section 179 kink point in the depreciation schedule. Standard errors clustered at the firm level are in parentheses.

we can use ineligible investment as an alternative intratemporal placebo test. The first two columns of Table 1.3 present two specifications of the intensive margin model, which replace eligible investment with structures investment. The first specification is the baseline model, and the second includes two-digit industry dummies interacted with quadratic time trends. We cannot distinguish the structures investment response from zero. Thus, the results pass this placebo test.

Another concern with our results is that they may merely reflect a reporting response, with much less actual investment taking place. The third and fourth columns of Table 1.3 provide a reality check. We replace our measure of investment derived from Form 4562 with net investment, which is the difference in logarithms of the capital stock between year t and year $t - 1$. Both the baseline and industry trend regressions confirm our gross investment results with net investment responding strongly as well.

Columns five and six of Table 1.3 offer a sanity check of our findings. Here, the dependent variable is an indicator for whether the firm reports depreciation expense in the specific form item applicable to bonus. Effectively, this is a test for bonus depreciation take-up. The table indicates that the probability of taking up bonus is strongly increasing in the strength of the incentive.

We present direct evidence that firms take the tax code into account when making investment decisions. With respect to equipment investment, they pay special attention to the depreciation schedule and the nonlinear incentives it creates. These nonlinear budget sets should induce *bunching* of firms at rate kinks. Consistent with this logic, we find sharp bunching at depreciation kink points. This evidence supports our claim that temporary bonus depreciation incentives were also salient.

To show policy salience, we study a component of the depreciation schedule, Section 179, which applies mainly to smaller firms. Under Section 179, taxpayers may elect to expense qualifying investment up to a specified limit. With the exception of used equipment,⁵³ all investment eligible for Section 179 expensing is eligible for bonus depreciation. Focusing on Section 179 thus serves as an out of sample test of policy salience that remains closely linked to the bonus incentives at the core of the paper.

Each tax year, there is a maximum deduction and a threshold over which Section 179 expensing is phased out dollar for dollar. The kink and phase-out regions have increased incrementally since 1993.⁵⁴ When the tax schedule contains kinks and the underlying distribution of types is relatively smooth, the empirical distribution should display excess mass at these kinks (Hausman, 1981; Saez, 2010). Figure 1.2 shows how dramatic the bunching behavior of eligible investment is in our setting. These figures plot frequencies of observations in our dataset for eligible investment grouped in \$250 bins. Each plot represents a year or group of years with the same maximum deduction, demarcated here by

⁵³Used equipment accounts for approximately six percent of equipment investment (Kitchen and Knittel, 2011).

⁵⁴Appendix Table A.1.1 summarizes the changes in Section 179 depreciation rules over the past twenty years.

a vertical line. The bunching within \$250 of the kink tracks the policy shifts in the schedule exactly and reflects a density five to fifteen times larger than the counterfactual distribution nearby.⁵⁵

In general, evidence of bunching at kink points reflects a mix of reporting and real responses.⁵⁶ The bunching evidence is informative in either case because these are both behavioral responses, which show whether firms understand and respond to the schedule. In the next section, we study managerial myopia by comparing bunching activity across different groups of firms. This test does not depend on whether the response is real or reported.

We can interact the bunching evidence with the basic regression model identifying the response to bonus. The design of the test generates control and treatment groups from the notion that firms differ in their tax code knowhow.⁵⁷ We compute geographic proxies of investment schedule sophistication through measuring the local propensity to bunch at the Section 179 kink point. We use the low information areas as cross-sectional counterfactuals for the high information areas. We then separately estimate the baseline model for each group, effectively providing a difference-in-difference-in-differences estimate of the bonus response.

⁵⁵Excess mass ratios are computed using the algorithm and code in Chetty et al. (2011).

⁵⁶See Saez (2010) for a discussion of this point. The bonus difference-in-differences (DD) design is less vulnerable to misreporting. In that design, we can confirm the response by looking at other outcomes. In addition, the DD estimator is much less sensitive to misreporting by a small fraction of total investment. Moreover, the sample contains many firms who use external auditors, for whom misreporting investment entails substantial risk and little benefit. Last, our conversations with tax preparers and corporate tax officers suggest that misreporting investment is an inferior way to avoid taxes. This is because investment purchases are typically easily verifiable, require receipts when audited, and usually reduce current taxable income by just a fraction of each dollar claimed as spent. In the case of investment expenses depreciated over multiple years, the audit risk of misreporting is also extended over the entire depreciation schedule.

⁵⁷This test follows the design of Chetty, Friedman and Saez (2013), who use geographic differences in individual bunching at a kink in the Earned Income Tax Credit schedule to study the labor supply response to taxes.

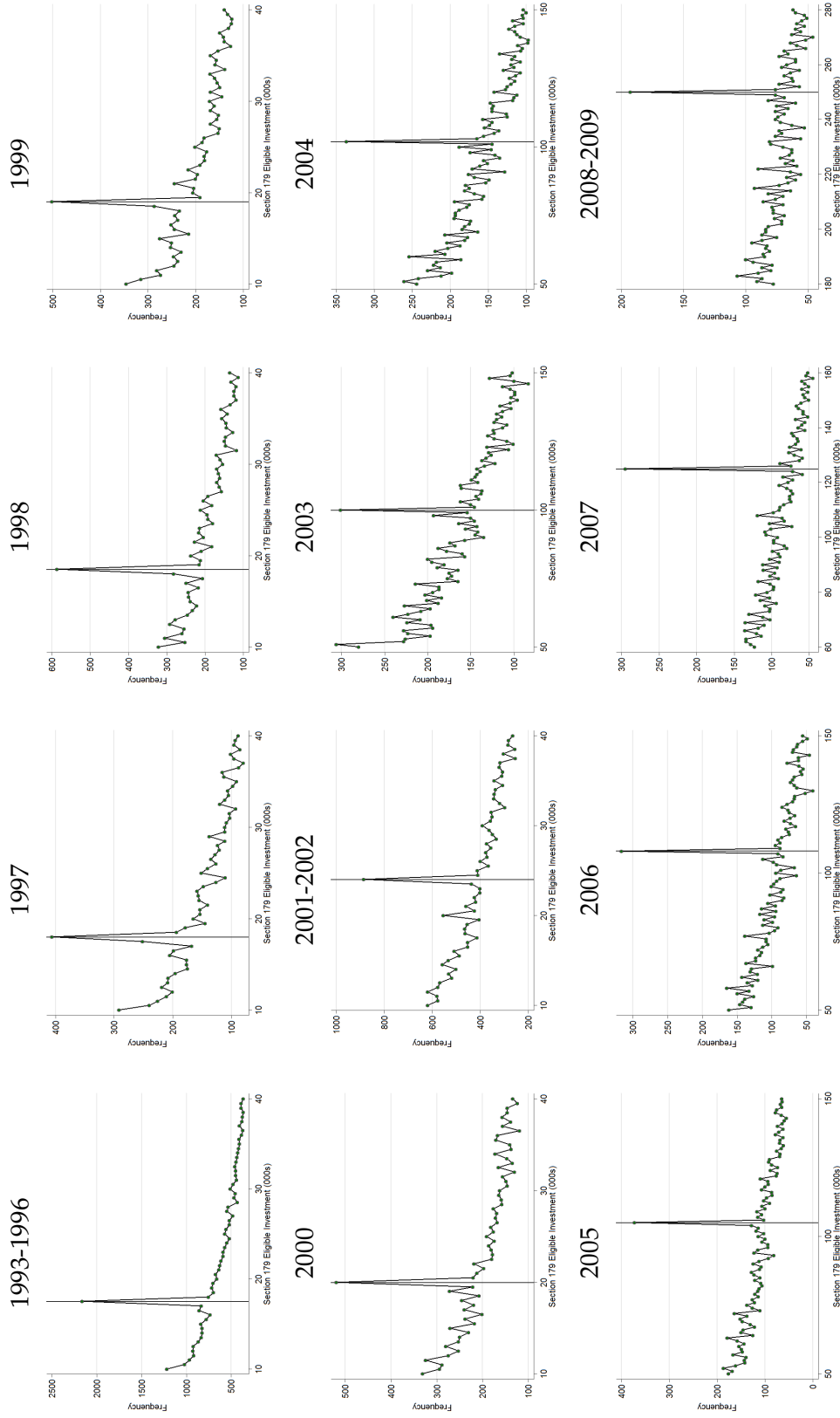


Figure 1.2: Depreciation Schedule Salience

Notes: These figures illustrate the salience of complex nonlinearities in the depreciation schedule. They show sharp bunching of Section 179 eligible investment around the depreciation schedule kink from 1993 through 2009. Each plot is a histogram of eligible investment in our sample in the region of the maximum deduction for a year or group of years. Each dot represents the number of firms in a \$250-bin. The vertical lines correspond to the kink point for that year or group of years. Bunching behavior by geography serves as a proxy for tax code sophistication.

We group firms by two-digit ZIP code, which is the lowest level of aggregation that permits a reliable measure of bunching. For each ZIP-2, we pool all years and compute the fraction of firms within \$10,000 of the kink who bunch within \$250 of it. This provides the sorting variable. In this design, more bunching in a region indicates more awareness of the tax code for that region. So, we should expect the growth in investment during bonus periods to be increasing in the level of bunching. Columns seven and eight of Table 1.3 show that indeed the high bunching areas display a stronger response to bonus than do the low bunching areas.⁵⁸

To recapitulate, bonus depreciation has a large effect on investment, and spurious time-varying industry factors cannot explain this fact. Such factors would cause parallel trends to fail in the years prior to bonus. They would lead to different estimates in recessions marked by weakness in different industries. They would lead ineligible investment to expand. They would attenuate the estimated effect when regressions include flexible industry-by-time controls. And they would lead to a similar response across geographies where firms pay more and less attention to the depreciation schedule. The facts do not match these predictions. Section 1.5, which presents heterogeneous effects by firm size, further contradicts the omitted industry factor story.

These investment responses directly correspond to take-up of depreciation incentives—bonus take-up rates rise with the policy’s generosity and many firms sharply bunch around the Section 179 kink point—in contrast to recent work on partial salience of sales taxes (Chetty, Looney and Kroft, 2009) and the nonresponse of investment to dividend tax changes (Yagan, 2013). Net investment responds to bonus depreciation as well, even though the reported balance sheet items do not affect taxable income. Section 1.6 shows that debt issuance increases because of bonus depreciation and that payroll and dividend payments—which are double reported—respond as well. Thus the observed response is a policy response that does not reflect a mere reporting response, but rather reflects real economic actions.

⁵⁸Specifically, we compare the top and bottom three deciles of local bunching.

1.5 Explaining the Large Response with Financial Frictions

The large response of investment to bonus depreciation is not consistent with a frictionless model of firm behavior: the magnitudes imply implausibly high discount rates. In this section, we explore alternative models that generate high effective discount rates.

One alternative is costly external finance, which raises the total discount rate firms apply to evaluate projects. Our rich data environment enables us to study how the investment response to tax incentives interacts with costly external finance. We perform a series of split sample tests, using several common markers of *ex ante* financial constraints.⁵⁹ Consistent with this story, firms more likely to depend on costly external finance—small firms, non-dividend payers and firms with low levels of cash—respond more strongly to bonus. Split sample analysis by size also leads to a more accurate prediction for the aggregate policy response.

Another alternative model is managerial myopia, which raises effective discount rates by sharply discounting the future relative to the present. Consistent with this story, firms only respond to investment incentives when the policy immediately generates after-tax cash flows. For firms with positive taxable income before depreciation, expanding investment reduces this year’s tax bill and brings extra cash into firm coffers today. Firms without this immediate incentive can still carry forward the deductions incurred but must wait to receive the tax benefits.⁶⁰ We present evidence that, for both Section 179 and bonus depreciation, this latter incentive is weak, and differences in growth opportunities cannot explain this fact.

⁵⁹See Fazzari, Hubbard and Petersen (1988a) for an early application of this methodology and Almeida, Campello and Weisbach (2004) and Chaney, Sraer and Thesmar (2012) for recent examples.

⁶⁰In the code, current loss firms have the option to “carry back” losses against past taxable income. The IRS then credits the firm with a tax refund. Our logic assumes that firms have limited loss carryback opportunities because, in the data, we find low take-up rates of carrybacks. Furthermore, carrybacks create a bias against our finding a difference between taxable and nontaxable firms, because carrybacks create immediate incentives for the nontaxable group.

1.5.1 Heterogeneous Responses by Ex Ante Financial Constraints

We divide the sample along several markers of ex ante financial constraints used elsewhere in the literature. Even for private unlisted firms, we can still measure size, payout frequency and proxies for balance sheet strength. Figure 1.3 plots elasticities and confidence bands from regressions run for each of ten deciles based on average sales.⁶¹ We plot both the intensive margin elasticities and the user cost elasticities. The smallest firms in the sample show the largest response to bonus, regardless of the specification. The user cost estimates help us reconcile our findings with those in past studies. Larger firms show user cost elasticities in line with the findings surveyed in Hassett and Hubbard (2002). It is only the smaller firms, for whom data were previously unavailable, that yield estimates outside the consensus range.

Table 1.4 presents a statistical test of the difference in elasticities across three markers of ex ante constraints. For the sales regressions, we split the sample into deciles based on average sales and compare the bottom three to the top three deciles.⁶² The average semi-elasticity for small firms is twice that for large firms and statistically significantly different with a p-value of 0.03.⁶³ The second two columns present separate estimates for firms who paid a dividend in any of the three years prior to the first round of bonus depreciation.⁶⁴ Here, the non-paying firms are significantly more responsive.

Our third sample split is based on whether firms enter the bonus period with relatively low levels of liquid assets. We run a regression of liquid assets on a ten-piece linear spline in total assets plus fixed effects for four-digit industry, time, and corporate form. We sort firm-year observations based on the residuals from this regression lagged by one year, and

⁶¹Specifically, we use average sales from the three years before each bonus period. We use as many of these six years as are available for each firm.

⁶²When we measure size with total assets or payroll, the size results are unchanged.

⁶³Cross equation tests are based on seemingly unrelated regressions with a variance-covariance matrix clustered at the firm level.

⁶⁴We only use the first round of bonus for the dividend split. The dividend tax cut of 2003, which had a strong effect on corporate payouts (Yagan, 2013), may have influenced the stability of this marker for the later period.

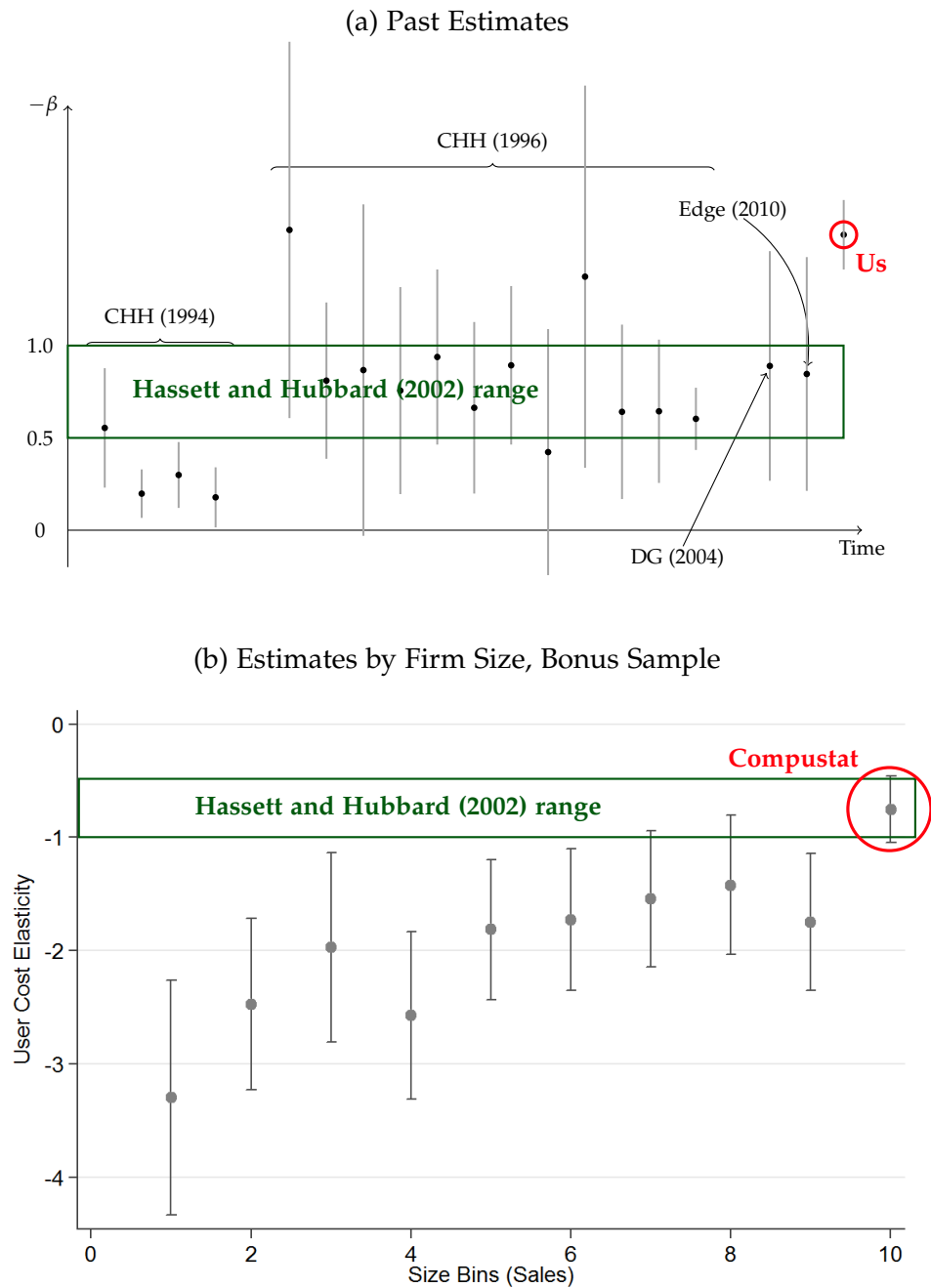


Figure 1.3: Heterogeneous Effects by Firm Size

Notes: These figures plot coefficients and confidence bands from user cost specifications (see the third row of Table 1.2) for past studies of tax reforms and our sample. The sources for the coefficients in Panel (a) are in Appendix Table A.1.2. Panel (b) splits the sample into deciles based on mean pre-policy sales. The average firm in Compustat during this time period falls in the tenth size bin (with sales equal to \$1.8B), which coincides with the Hassett and Hubbard (2002) survey range of user cost elasticity estimates (-0.5 to -1).

then report in the last two columns of Table 1.4 separate estimates for the top and bottom three deciles. These are reported in the last two columns of Table 1.4. The results using this marker of liquidity parallel those in the size and dividend tests, with the low liquidity firms yielding an estimate of 7.2 as compared to 2.8 for the high liquidity firms.

Table 1.4: Heterogeneity by Ex Ante Constraints

	Sales		Div Payer?		Lagged Cash	
	Small	Big	No	Yes	Low	High
$z_{N,t}$	6.29*** (1.21)	3.22*** (0.76)	5.98*** (0.88)	3.67*** (0.97)	7.21*** (1.38)	2.76** (0.88)
Equality Test	$P = .030$		$P = .079$		$P = .000$	
Observations	177620	255266	274809	127523	176893	180933
Clusters (Firms)	29618	29637	39195	12543	45824	48936
R ²	0.44	0.76	0.69	0.80	0.81	0.76
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table estimates regressions from the baseline intensive margin specification presented in Table 1.2. Here, we split the sample based on pre-policy markers of financial constraints. For the size splits, we divide the sample into deciles based on the mean value of revenues, with the mean taken over years 1998 through 2000. Small firms fall into the bottom three deciles and big firms fall into the top three deciles. For the dividend payer split, we divide the sample based on whether the firm paid a dividend in any of the three years from 1998 through 2000. The dividend split only includes C corporations. The lagged cash split is based on lagged residuals from a regression of liquid assets on a ten piece spline in total assets and fixed effects for four-digit industry, year and corporate form. The comparison is between the top three and bottom three deciles of these lagged residuals. Standard errors clustered at the firm level are in parentheses.

These constraint markers are imperfect.⁶⁵ First, they do not directly measure the external finance cost faced by new firms. This concern would tend to bias any differences existing between groups toward zero, and thus against the results we present. A second concern with sample splitting is that the splitting criteria are correlated with the investment error term and so may bias the estimated coefficient of interest (Bond and Van Reenen, 2007). This issue is important for investment-cash flow sensitivity tests because cash flow is likely correlated with other components of the investment error term. Because our setting features plausibly exogenous policy variation at the industry level, this concern is less important

⁶⁵Criticism of split sample markers dates back to Poterba's comments in Fazzari, Hubbard and Petersen (1988a). See Farre-Mensa and Ljungqvist (2013) for a more recent assessment of their value in samples of public and private companies.

here. The key assumption we make is that interacting our splitting criterion, measured prior to the policy change, with the policy variable and the year effects enables a valid difference-in-differences design for each group.

1.5.2 Heterogeneous Responses by Tax Position

The Section 179 bunching environment presents an elegant setting to document the immediacy of investment responses to tax policy. The simple idea is to separate firms based on whether their investment decisions will fully offset current year taxable income, or whether deductions will have to be carried forward to future years. We choose net income before depreciation expense as our sorting variable. Firms for which this variable is positive have an immediate incentive to invest and reduce their current tax bill. If firms for which this variable is negative show an attenuated investment response and these groups are sufficiently similar, we can infer that the immediate benefit accounts for this difference.

The panels of Figure 1.4 starkly confirm our intuition. In panel (a), we pool all years in the sample, recenter eligible investment around the year's respective kink, and split the sample according to a firm's taxable status. Firms in the left graph have positive net income before depreciation and firms in the right graph have negative net income before depreciation. For firms below the kink on the left, a dollar of Section 179 spending reduces taxable income by a dollar in the current year. Retiming investment from the beginning of next fiscal year to the end of the current fiscal year can have a large and immediate effect on the firm's tax liability. For firms below the kink on the right, the incentive is weaker because the deduction only adds to current year losses, deferring recognition of this deduction until future profitable years. As the figure demonstrates, firms with the immediate incentive to bunch do so dramatically, while firms with the weaker, forward-looking incentive do not bunch at all.

One objection to the taxable versus nontaxable split is that nontaxable firms have poor growth opportunities and so are not comparable to taxable firms. We address this objection in two ways. First, we restrict the sample to firms very near the zero net income before

depreciation threshold to see whether the difference persists when we exclude firms with large losses. Panel (a) of Figure A.1.1 plots bunch ratios for taxable and nontaxable firms, estimated within a narrow bandwidth of the tax status threshold. The difference in bunching appears almost immediately away from zero, with the confidence bands separating after we include firms within \$50 thousand dollars of the threshold. For loss firms, the observed pattern cannot be distinguished from a smooth distribution, even for firms very close to positive tax position. The bunching difference for nontaxable firms is not driven by firms making very large losses.

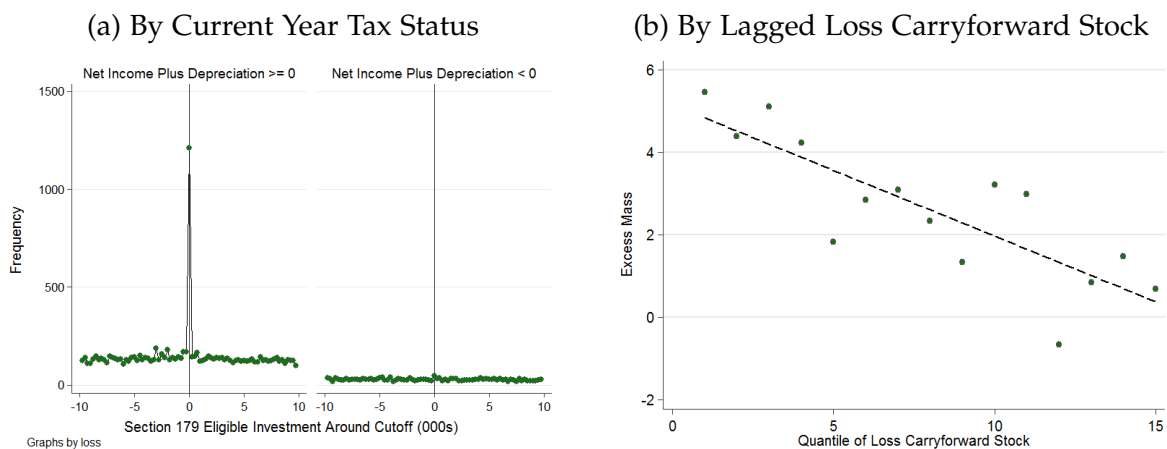


Figure 1.4: *Bunching Behavior and Tax Incentives*

Notes: These figures illustrate how bunching behavior responds to tax incentives. Firms bunch less when eligible investment provides less cash back now. Panel (a) splits the sample based on whether firm net income before depreciation is greater than or less than zero. Firms with net income before depreciation less than zero can carry back or forward deductions from eligible investment but have no more current taxable income to shield. Panel (b) groups firms with current year taxable income based on the size of their prior loss carryforward stocks. The x-axis measures increasing loss carryforward stocks relative to current year income. The y-axis measures the excess mass at the kink point for that group. Firms with more alternative tax shields find investment a less useful tax shield and therefore bunch less.

Table 1.5 replicates the tax status split idea in the context of bonus depreciation. We modify the intensive margin model from Table 1.2 by interacting all variables with a taxable indicator based on whether net income before depreciation is positive or negative.⁶⁶ According to these regressions and consistent with bunching results, the positive effect of

⁶⁶That is, we interact z , any controls, and the time fixed effects with the taxable indicator. We do not interact the firm effects with the taxable indicator.

bonus depreciation on investment is concentrated exclusively among taxable firms. The semi-elasticity is statistically indistinguishable from zero for nontaxable firms, while it is 3.8 for taxable firms. In panel (b) of Figure A.1.1, we repeat the narrow bandwidth test for bonus depreciation. The figure plots the coefficients on the interaction of taxable and nontaxable status with the policy variable. The difference in coefficients in Table 1.5 emerges within \$50 thousand of the tax status threshold, and these coefficients are statistically distinguishable within \$100 thousand of the threshold. Here as well, the results are not driven by differences for firms far from positive tax positions.

Table 1.5: Heterogeneity by Tax Position

	LHS Variable is Log(Investment)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Taxable	3.83***	3.08***	1.95*	6.43***	4.32***	4.15***	
× $z_{N,t}$	(0.79)	(0.93)	(0.92)	(1.46)	(0.96)	(0.82)	
$z_{N,t}$	-0.15	0.60	0.38	-3.03*	-0.69	0.88	5.68***
	(0.90)	(1.05)	(1.06)	(1.55)	(1.15)	(0.94)	(1.70)
Medium LCF							-2.56
× $z_{N,t}$							(1.46)
High LCF							-3.70*
× $z_{N,t}$							(1.55)
$CF_{it}/K_{i,t-1}$		0.14***					
		(0.028)					
Taxable		0.27***					
× $CF_{it}/K_{i,t-1}$		(0.035)					
Observations	735341	580422	514035	221306	585914	722262	119628
Clusters (Firms)	128001	100883	109678	63699	107985	124962	40282
R ²	0.71	0.74	0.74	0.80	0.73	0.72	0.84
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	Yes	No	No
Industry Trends	No	No	No	No	No	Yes	No

Notes: This table estimates regressions from each intensive margin in columns (1) through (6) specification presented in Table 1.2. For each firm year, we generate an indicator based on whether a firm is in taxable position prior to depreciation expense. We fully interact this indicator with all controls and the time effects. Column (7) splits taxable firms into three groups based on the size of their lagged loss carryforward stocks relative to net income before depreciation. We interact these group indicators with $z_{N,t}$ and the time effects. Only firms with nonzero stocks of lagged loss carryforwards are included. Standard errors clustered at the firm level are in parentheses.

To address the concern about nontaxable firms, Panel (b) of Figure 1.4 uses differences

within the group of taxable firms. This plot shows again that bunching is due to tax planning with regard to the immediate potential benefit. Here, we divide profitable firms by their stock of loss carryforwards in the previous year. Each dot in this plot represents a bunching histogram where the y-axis measures the degree of bunching using the excess mass estimator in Chetty et al. (2011). The groups are sorted according to the ratio of lagged loss carryforward stock to current year net income before depreciation, which proxies for the availability of alternative tax shields. The scatter clearly indicates a negative relationship between the presence of this alternative tax shield and the extent of eligible investment manipulation.

We confirm this pattern in the bonus setting. Column (7) of Table 1.5 focuses on the group of taxable firms with non-zero stocks of lagged loss carryforwards. We split this group into three subgroups based on the size of their carryforward stock. Firms with large stocks of loss carryforwards display a semi-elasticity with respect to z of 2 compared to a semi-elasticity of 5.7 for firms with low loss carryforward stocks.

The finding for nontaxable firms contradicts a simple model of costly external finance, because firms neglect how the policy affects borrowing in the future. On the other hand, firms cannot be too myopic because the investment decision itself only pays off in the future. Thus for myopia to be the explanation, firms must use different accounts to think about investment decisions and the tax implications. Moreover, the myopia story needs complexity to explain the finding for financially constrained firms—are small firms, non-dividend payers and firms with low levels of cash more myopic? While plausible, there is no evidence of this.

The facts presented in this section—the stronger response for financially constrained firms and the nonresponse for nontaxable firms—do not match the predictions of a frictionless model. The facts point instead toward models in which costly external finance matters and current benefits outweigh future benefits, with neither alternative being obviously redundant.

In the next section, we use an investment model and the estimates to calibrate a parameter

for each alternative model.

1.5.3 Discount Rates and the Shadow Cost of Funds

Taken together, our empirical findings emphasize a financial frictions channel for how investment incentives work. We use a standard investment model to quantify the importance of this channel. Specifically, we ask what is the marginal value of cash, λ , implied by our financial constraint split sample analysis, and what is the discount term, β , implied by our tax status split sample analysis. The answers combine to tell us what discount rates firms apply when making investment decisions.

In Appendix A.1, we derive the comparative static for investment with respect to the bonus depreciation term θ :

$$I \cdot \varepsilon_{I,\theta} \equiv \frac{\partial I}{\partial \theta} = \frac{(1 + \lambda)p_I}{\psi_{II}} \frac{\partial z}{\partial \theta} > 0, \quad (1.8)$$

where $\varepsilon_{I,\theta}$ is the semi-elasticity of investment with respect to θ , p_I is the price of investment, ψ_{II} is the second derivative of the adjustment cost function, and z is defined as in (1.3). In the Appendix, we state assumptions under which $I \cdot \psi_{II}$ will be equal across groups.⁶⁷ Under these assumptions, we can derive two empirical moments that combine our estimates for constrained and unconstrained firms and for taxable and nontaxable firms and yield simple formulas for λ and β .

The first empirical moment we use compares the estimated response with respect to bonus for constrained and unconstrained firms. Assuming constrained firms face shadow price λ_C and unconstrained firms face shadow price λ_U , we take the ratio of comparative statics:

$$\frac{\varepsilon_{I,z}^C}{\varepsilon_{I,z}^U} \equiv m_1 = \frac{\partial I / \partial \theta|_{\lambda_C}}{\partial I / \partial \theta|_{\lambda_U}} = \frac{1 + \lambda_C}{1 + \lambda_U} = 1 + \frac{\Delta \lambda}{1 + \lambda_U}, \quad (1.9)$$

which reveals an implied credit spread between constrained and unconstrained firms. Table 1.6 presents m_1 for each pair of estimates in Table 1.4. λ is the shadow price of relaxing the

⁶⁷That is, we assume linear homogeneity of the marginal adjustment cost function. Nearly all studies in the literature make this assumption, which is necessary for example for marginal q to equal average Q .

Table 1.6: Calibrated Moments

Shadow Cost of Funds Calibration				
	Mean Sales	Dividend Payers	Lagged Cash	Average
m_1	1.95	1.63	2.61	2.06
$\lambda_C _{\lambda_U=0}$	0.95	0.63	1.61	1.06
Discount Factor Calibration				
	p	ϕ	r	β
High ϕ	.5	.88	.07	0.84
Low ϕ	.1	.59	.07	0.92

Notes: This table computes empirical estimates for m_1 and m_2 , as defined in the text. m_1 reveals an implied credit spread between constrained and unconstrained firms. m_2 reveals the discount factor firms apply to all future cash flows relative to current flows.

firm's borrowing constraint. An alternative interpretation is that every after-tax dollar inside the firm is worth $1 + \lambda$ dollars outside the firm. Our estimates reveal that, for financially constrained firms, a dollar inside the firm is worth \$2.06 on average outside the firm.

Is this estimate reasonable? There are not many existing benchmarks. Faulkender and Wang (2006) attempt a calculation with a very different methodology, but that ultimately arrives at a similar conclusion. They estimate the value of changes in cash in excess return regressions, while attempting to control for a host of omitted factors. They find that for low payout firms and for small firms the value of a dollar of after-tax cash is worth \$1.67 and \$1.62, respectively. For these firms' unconstrained counterparts, a dollar is only worth \$1.07 and \$1.12. The spreads in their study are comparable to ours, especially considering their exercise operates within a group of firms we consider to be relatively unconstrained.⁶⁸

We define a second empirical moment that compares taxable and nontaxable firms:

$$\frac{\varepsilon_{I,z}^{\gamma=0}}{\varepsilon_{I,z}^{\gamma=1}} \equiv m_2 = \frac{\partial I / \partial \theta|_{\gamma=0}}{\partial I / \partial \theta|_{\gamma=1}} = \beta \phi \frac{1 - z_t^0(1)}{1 - z_t^0(\beta)}, \quad (1.10)$$

where ϕ is a discount factor that reflects the average arrival of the taxable status event for

⁶⁸Similarly, Kojien and Yogo (2012) find that a relaxed borrowing constraint for life insurers is worth \$2.32 per dollar of inside capital.

nontaxable firms. We proxy for ϕ by assuming a fixed transition probability p for nontaxable firms and an infinite horizon for carryforward realization.⁶⁹ This implies $\phi = p/(p + r)$.⁷⁰ We calibrate the transition probability using the probability that an actually nontaxable firm transitions into tax status in the next period. In our data, this probability is approximately 0.5.⁷¹ The sample we compare restricts the nontaxable group to include only firms that are nearly taxable, so this p offers a conservative estimate of ϕ . We compute the implied β for this p and for an extreme p equal to 0.1.

Note the external finance wedge falls out of this expression. This is true as long as average shadow costs are the same across taxable and nontaxable groups. To maintain this assumption, we use our loss carryforward group estimates to calibrate m_2 . That is, we estimate semi-elasticities within the group of taxable firms sorted according to their past stocks of alternative tax shields. For firms with large loss carryforward stocks relative to current income, the marginal dollar of investment is unlikely to affect this year's tax bill. At the same time, we have less reason to believe these firms face substantially worse growth opportunities or tighter financial constraints. This biases our estimates of β toward the neoclassical benchmark of β equal to one.

Applying the estimates from the last column of Table 1.5 yields a value for m_2 of 0.35(= (5.68 – 3.7)/5.68). For $p = 0.5$, this maps to an implied discount factor (β) of 0.84. Ignoring for the moment the other discount terms, β equal to 0.84 implies a discount rate of approximately 17 percent. We are not aware of studies that attempt to measure discount factors such as this for firms. Prior studies on individual decision making have found similar magnitudes for short term discount rates in both lab and field experiments.⁷²

⁶⁹The actual expiration period for carryforwards is twenty years.

⁷⁰That is, the expected arrival is $p/(1 + r) + (1 - p)p(1 + r)^{-2} + (1 - p)^2p(1 + r)^{-3} + \dots = p/(1 + r) \cdot [1/(1 - (1 - p)/(1 + r))] = p/(p + r)$.

⁷¹Auerbach and Poterba (1987) note more persistence of nontaxable positions than we do. Our measure is based on net income before depreciation, in order to capture the state of having the next dollar of investment affect this year's tax bill. Their measure is based on whether firms exhaust their carryforward stocks. Below, we assess the robustness of our results to varying ϕ .

⁷²Laibson D. Repetto and Tobacman (2007) estimate short term discount rates of 40 percent in the context of individual saving decisions. In a more general model, they estimate a short run discount rate of 15 percent and

The discounting implied by β says that one dollar next year is worth 84 cents, before taking into account risk or the shadow cost of funds. If we then apply the assumed risk adjusted rate of 7 percent and the estimated shadow cost of funds of 1.06, we find that a dollar next year is worth approximately 38 cents today for the credit constrained firms in our sample. This substantial discount is not surprising, given the starkness of the reduced form empirical results: nontaxable firms seem to ignore the future benefits and small, financially constrained firms seem to value highly the immediate cash back due to bonus depreciation. In the model, we use costly external finance and myopia to describe the deviations from a rational benchmark we observe, but the exercise performed here provides just one of several plausible calibrations of this basic fact. In general, models of firm behavior that do not generate high discount rates are unlikely to fit the data for most firms.

1.6 Substitution Margins and External Finance

We ask whether increased investment involves substitution away from payroll or equipment rentals, how firms finance their additional investment, and whether the increased investment reflects intertemporal substitution or new investment. Studying external finance responses helps us understand how firms paid for new investments. Understanding substitution margins is critical for assessing the macroeconomic impact of these policies.

Table 1.7 presents estimates of the intratemporal and intertemporal substitution margins. These regressions follow the baseline specification in equation 1.7, with a different left hand side variable. For rents, payroll and debt, we focus on flows⁷³ as outcomes that match investment (i.e., capital stock flows) most closely. For payouts, we study the logarithm of dividend payments and an indicator for whether dividends are non-zero, which separates the intensive and extensive margins.

How flexible is the rent-versus-own margin for equipment investment? This is a crucial

a long run rate of 3 percent.

⁷³Namely, differences in logs.

question for assessing the real effect of these stimulus policies. If firms simply shift away from leasing to take advantage of the tax benefits of buying, then the aggregate impact of these policies will be minimal. In their tax returns, firms separately report rental payments for computing net income. Unfortunately, this item does not permit decomposition into equipment and structures leasing. Given this limitation, we can still ask what effect bonus depreciation had on changes in rental payments. The first column of Table 1.7 shows that growth in rental payments did not slow due to bonus, but rather increased somewhat. Thus, we do not find evidence of substitution away from equipment leasing. The second column of Table 1.7 reports the effect of bonus on growth in non-officer payrolls. Again, we find no evidence of substitution, but rather coincident growth of payroll. Finding limited substitution in both leasing and employment makes it more likely that bonus incentives caused more output.

While increased depreciation deductions do allow firms to reduce their tax bills and keep

Table 1.7: Substitution Margins and External Finance

	Dependent Variable					
	Δ Rents	Δ Payroll	Δ Debt	Dividends	Payer?	Investment
	(1)	(2)	(3)	(4)	(5)	(6)
$z_{N,t}$	0.75** (0.26)	1.49*** (0.20)	1.84*** (0.21)	-2.14*** (0.54)	-0.36*** (0.089)	4.22*** (0.62)
$z_{N,t-2}$						-0.86 (0.69)
Observations	574305	624918	642546	133161	818576	476734
Clusters (Firms)	98443	102043	103868	28891	128150	84777
R ²	0.18	0.23	0.20	0.90	0.68	0.76
Year Effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table estimates regressions of the form

$$X_{it} = \alpha_i + \beta z_{N,t} + X_{it} + \delta_t + \varepsilon_{it}$$

where X_{it} equals the difference in the logarithm of the dependent variable in columns (1) through (3). In columns (4) and (5), the dependent variable is the logarithm of dividends paid or an indicator for positive dividend payments. $z_{N,t}$ is the present value of a dollar of eligible investment computed at the four-digit NAICS industry level, taking into account periods of bonus depreciation. Column (6) includes contemporaneous and twice lagged $z_{N,t}$. Standard errors clustered at the firm level are in parentheses.

more cash inside the firm, they must still raise adequate financing to make the purchases in the first place. This point is especially critical if, as the data suggest, firms thought to be in tight financial positions respond more. Here, we test whether bonus incentives affect net issuance of debt and payout policy. Columns three through five of Table 1.7 provide some insight. Increased equipment investment appears to coincide with significantly expanded borrowing and reduced payouts both on the extensive and intensive margins.

We assess the extent of intertemporal substitution using a model that includes both contemporaneous z and lagged z . Our data often do not include the fiscal year month, so it is possible that we are marking some years as t when they should be $t - 1$ or $t + 1$. For most of our tests, this issue introduces an attenuation but no systematic bias. However, when testing for intertemporal substitution, we want to be sure that lagged z measures past policy changes. Thus, column (6) of Table 1.7 includes regressions with twice lagged z added to the baseline bonus model. The coefficient on lagged z is negative but not distinguishable from zero and including lagged z does not alter the coefficient on contemporaneous z . This implies limited intertemporal shifting of investment.

1.7 Conclusion

This paper combines methods from public and applied economics with insights from finance to answer a first order macroeconomic question: how do taxes affect investment behavior in the presence of financial frictions? We find that firms respond strongly to incentives that directly target investment decisions. Our heterogeneity results—that the investment response is larger for financially constrained firms, but only when the benefit is immediate—show that financial frictions are critical for understanding investment behavior.

The results point toward a set of models in which costly external finance matters and firms place more weight on current benefits than they would in a frictionless model. Whether the high implied discount rate reflects an external finance wedge, managerial myopia, agency considerations or a mix of these is an important question for future research. Further study of the external finance mechanism would be valuable. A deeper study of the

employment effects of these policies is of direct interest to macroeconomic modelers.

A related question for future research concerns the effects of tax planning. How do tax preparers affect the decision to take up these policies? More generally, do firms focus on minimizing current taxes at the possible expense of future payoffs? The answer to these questions might shed light on the role of agency problems and firm learning about optimal management practices.

The empirical results imply that policies which target investment directly and yield immediate payoffs are most likely to influence investment activity. Policies that target financial constraints, such as direct loans, might have a similar effect if conditional on the investment decision. In comparison to studies of consumer durable goods, we find less evidence of intertemporal shifting, but more work on this question is needed. Data from the period following the recent stimulus, once available, will be very useful.

Chapter 2

Arrested Development: Theory and Evidence of Supply-Side Speculation in the Housing Market¹

2.1 Introduction

How do prices aggregate information? We take up this question in a setting of particular macroeconomic importance: housing markets. Housing is a key driver of the business cycle (Leamer, 2007), and the causes of the financial crisis of 2008 and the Great Recession originated in housing markets (Mian and Sufi, 2009, 2011). An enduring feature of these markets is booms and busts in prices that coincide with widespread disagreement about fundamentals (Shiller, 2005). This paper argues that these cycles are caused by how housing markets aggregate beliefs.

Studying belief aggregation allows us to address some of the most puzzling aspects of the U.S. housing boom that occurred between 2000 and 2006. According to the standard model of housing markets, elastic housing supply prevents house price booms by allowing

¹Co-authored with Charles G. Nathanson

new construction to absorb rising demand.² But the episode from 2000 to 2006 witnessed several major anomalies, in which historically elastic cities experienced house price booms despite continuing to build housing rapidly. And house prices rose more in many of these cities—located in Arizona, Nevada, inland California, and Florida—than in cities where it was difficult to build new housing. Further complicating the puzzle, house prices remained flat in other elastic cities that were also rapidly building housing. Why was rapid construction able to hold down house prices in some cities and not others?

We solve this puzzle by adding two ingredients to the standard model. The first is a friction that makes owner-occupancy more efficient than renting. The second is disagreement about long-run growth paths. In this framework the way housing markets aggregate beliefs depends on a city's land availability. Prices appear more optimistic when land is plentiful and building houses is easy, reversing the standard model's intuition for how land supply influences prices. Crucially, optimism amplifies prices most when a city nears but has not yet reached a long-run development constraint. This mechanism matches the data. The anomalous cities are those that, as the boom began, found themselves in just this state of "arrested development."

We model a city of developers and residents with a fixed amount of land available for development. Developers decide how many houses to build and how much land to buy. Residents decide how much housing to consume and whether to buy or rent. They prefer owning their houses over renting because of frictions in the rental market.³ Residents can invest in the equity of developers, which provides exposure to land prices. Short-selling land and housing is impossible, but residents can short-sell developer equity. Over time, new residents arrive in the city, leading developers to build houses using their holdings of undeveloped land. Because of this growth, the city gradually exhausts its land supply. What today's investors believe about future inflows determines the price of undeveloped land.

²See, for example, Glaeser, Gyourko and Saiz (2008), Gyourko (2009), and Saiz (2010).

³Such frictions include the effort spent monitoring tenants to prevent property damage (Henderson and Ioannides, 1983), tax disadvantages (Poterba, 1984), and difficulty renting properties like single-family homes that are designed for owners (Glaeser and Gyourko, 2009).

House construction is instantaneous and developers bear a constant unit cost per house. As a result, all variation in house prices is caused by movements in land prices and not construction costs. Data from the U.S. boom support this feature of the model. Rapidly rising land prices account for most of the house price increases across cities. In contrast, construction costs remained relatively stable throughout the boom, and cost changes hardly varied across cities. These aspects of the data distinguish our theory from those that stress “time-to-build” factors such as input shortages or delivery lags (Mayer and Somerville, 2000; Gao, 2014).

We study a demand shock that raises the current inflow of new residents and also creates uncertainty about future inflows. Disagreement about long-run demand leads to disagreement about future house prices. The most optimistic residents seek to speculate through buying housing and through buying the equity of optimistic developers who are buying land.

Our first result is that speculation is crowded out of the housing market and into the land market. Consider an optimistic resident who wishes to speculate on future house prices. Buying a house and renting it out is difficult because of the widespread preference for owner-occupancy. And buying more housing for personal consumption is unappealing because of diminishing marginal utility. Land however offers a pure, frictionless bet on real estate. The optimistic resident chooses to invest in land through buying developer equity.

With data from the U.S. housing boom, we confirm several of the model’s predictions about land speculation. In the model, developers run by optimistic CEOs use resident financing to amass large land portfolios, buying land from less optimistic developers. Consistent with this prediction, we find that supply-side speculation figures prominently in the data. Between 2000 and 2006, the eight largest U.S. public homebuilders tripled their land investments, an increase far exceeding their additional construction needs. Their market equity then fell 74%, with most of the losses coming from write-downs on their land portfolios. The model also predicts that short-selling of developer equity increases during a boom because pessimistic residents disagree with the high valuations of the developer land

portfolios. Matching this prediction, the short interest in homebuilder stocks rose from 2% in 2001 to 12% in 2006. Rising short interest provides evidence of disagreement over the value of homebuilder land portfolios and thus over future house prices.

Our second result concerns how house prices aggregate beliefs. Speculators are crowded into the land market, while homeownership remains dispersed among residents of all beliefs. Therefore, house prices reflect a weighted average of the optimistic belief of speculators and the average owner-occupant belief. The weight on the optimistic belief equals the share of the housing market on the margin that consists of the land market. Prices look most optimistic where land is plentiful and building easy—that is, in cities where the short-run elasticity of housing supply is large.

This optimism bias affects prices most when the city's housing supply will become inelastic soon. This observation, which constitutes our third result, explains why house price booms occur in some elastic cities and not others. Consider a city in which the land available for development is large relative to the city's current size. Here, new construction fully absorbs the demand shock now and in the foreseeable future, and so beliefs about future house prices remain unchanged. The shock raises future price expectations only in cities where construction will be difficult in the near future.

Speculation amplifies house price booms most in cities that exist in a state of arrested development: they have ample land for construction today, but also face land barriers that will restrict growth in the near future. This theoretical supply condition characterizes the anomalous elastic cities during the U.S. housing boom. For instance, Las Vegas faces a development boundary put in place by Congress in 1998 and depicted in Figure 2.1. During the 2000-2006 housing boom, many investors believed the city would soon run out of land.⁴ Likewise, Phoenix's long-run development is constrained by Indian reservations

⁴Las Vegas provides a particularly clear illustration of our model. The ample raw land available in the short-run allowed Las Vegas to build more houses per capita than any other large city in the U.S during the boom. At the same time, speculation in the land markets caused land prices to quadruple between 2000 and 2006, rising from \$150,000 per acre to \$650,000 per acre, and then lose those gains. This in turn led to a boom and bust in house prices. The high price of \$150,000 for desert land before the boom and after the bust demonstrates the binding nature of the city's long-run development constraint. A *New York Times* article published in 2007 cites investors who believed the remaining land would be fully developed by 2017 (McKinley and Palmer, 2007).

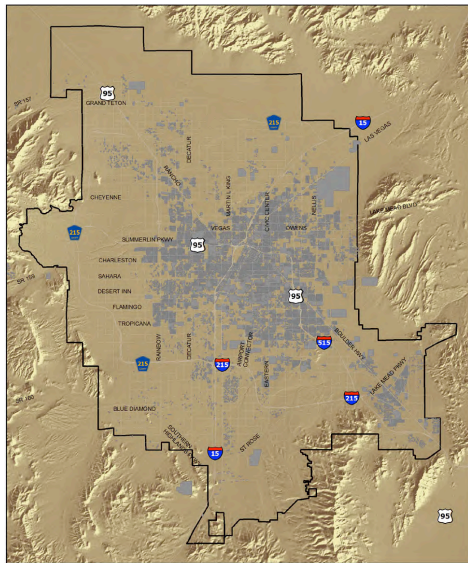
and National Forests that surround the metropolitan area (Land Advisors, 2010). In inland California, much of the farmland around cities is protected by a state law that penalizes real estate development on these parcels (Onsted, 2009).

When disagreement is strong enough, house prices increase more in these nearly developed cities than in a fully developed city. In the nearly developed cities, the extreme optimistic beliefs of land speculators determine house prices, amplifying the house price boom. Prices remain more stable in the fully developed city because they reflect the average belief. This result explains the puzzling house price booms in elastic areas that motivate this paper. Supply conditions in these places—elastic current supply, inelastic long-run supply—lead disagreement to have the largest possible amplification effect on a house price boom.

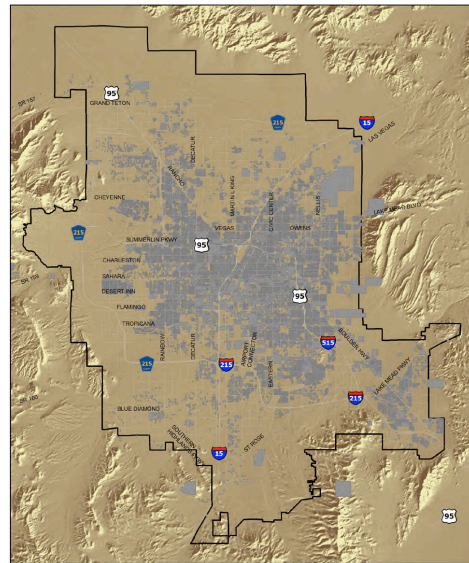
Our theory differs from several other explanations for the strong house price booms that occurred in elastic areas between 2000 and 2006. One possibility is that these cities experienced much larger demand shocks than the rest of the United States.⁵ Our analysis assumes a constant demand shock across cities; the heterogeneity in city house prices booms results entirely from differences in supply conditions. An additional possibility is that uncertainty increased land values due to the embedded option to develop land with different types of housing (Titman, 1983; Grenadier, 1996), and that this option value increase was largest in cities with an intermediate amount of land. In our model, all housing is identical, so this option does not exist. A final explanation is that developers hoarded land to gain monopoly power, and the incentive to do so was strongest in cities about to run out of land. This effect does not appear in our model because homebuilding is perfectly

The dramatic rise in land prices during the boom resulted from optimistic developers taking large positions in the land market. In a striking example of supply-side speculation, a single land development fund, Focus Property Group, outbid all other firms in every large parcel land auction between 2001 and 2005 conducted by the federal government in Las Vegas, obtaining a 5% stake in the undeveloped land within the barrier. Focus Property Group declared bankruptcy in 2009.

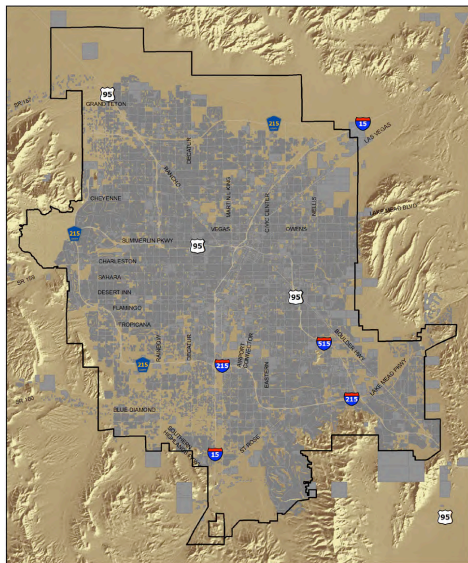
⁵For instance, the expansion of credit described by Mian and Sufi (2009) may have been largest in these cities. Alternatively, historical increases in house prices in nearby areas may have spread to these cities, either through behavioral contagion (DeFusco et al., 2013) or long-distance gentrification (Guerrieri, Hartley and Hurst, 2013).



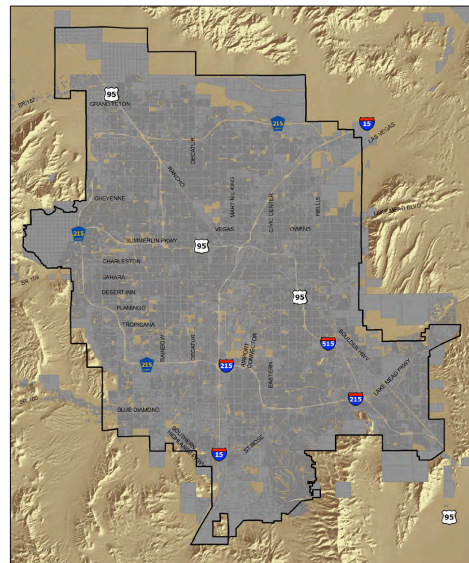
1980



1990



2008



2030

Figure 2.1: Long-Run Development Constraints in Las Vegas

Notes: This figure comes from Page 51 of the Regional Transportation Commission of Southern Nevada's Regional Transportation Plan 2009-2035 (RTCSNV, 2012). The first three pictures display the Las Vegas metropolitan area in 1980, 1990, and 2008. The final picture represents the Regional Transportation Commission's forecast for 2030. The boundary is the development barrier stipulated by the Southern Nevada Public Land Management Act. The shaded gray region denotes developed land.

competitive, as is the case empirically at the metro-area level.⁶ Unlike these stories, our approach explores the cross-sectional implications of disagreement, an understudied aspect of housing cycles for which we provide direct evidence.

In addition to explaining the city-level cross-section, our model offers new predictions on the cross-section of neighborhoods within a city. We allow some residents to prefer renting over owner-occupancy, so that both rental and owner-occupied housing exist in equilibrium. During periods of disagreement, optimistic speculators hold the rental housing, just as they hold land. Prices appear more optimistic, and hence house price booms are larger, in neighborhoods where a greater share of housing is rented. This prediction matches the data: house prices increased more from 2000 to 2006 in neighborhoods where the share of rental housing in 2000 was higher.

A long literature in macroeconomics and finance has studied how prices aggregate information. When markets are complete and investors share a common prior, prices usually are efficient and reflect the information of all market participants (Fama, 1970; Grossman, 1976; Hellwig, 1980). Our paper sits among a body of work showing that prices reflect only a limited and potentially biased subset of information when investors persistently disagree with each other, and markets are incomplete. Many of these papers focus on strategic considerations that arise in this setting, and the implications for asset prices (Harrison and Kreps, 1978; Scheinkman and Xiong, 2003). A related literature, starting with Miller (1977), demonstrates that prices can be biased even in the absence of strategic considerations because optimists end up holding the asset.⁷ We show that this optimism

⁶Somerville (1999) demonstrates the high level of homebuilder competition at the metro-area level, although he points out that construction is less competitive at the neighborhood level. Hoberg and Phillips (2010) argue that price booms often occur in competitive industries because firms mistakenly believe they will obtain future monopoly power.

⁷In these papers, all market participants are fundamental investors who ignore other investors' beliefs (Chen, Hong and Stein, 2002; Geanakoplos, 2009; Hong and Sraer, 2012; Simsek, 2013*a,b*). Pástor and Veronesi (2003, 2009) also study environments in which investors care only about long-run fundamentals during booms and busts, but their focus is on learning, and all investors agree as they are all identical. Piazzesi and Schneider (2009) and Burnside, Eichenbaum and Rebelo (2013) also apply models of disagreement to the housing market. Papers in which strategic behavior matters include Abreu and Brunnermeier (2003), Allen, Morris and Shin (2006), and Hong, Scheinkman and Xiong (2006).

bias is strongest in housing markets when land is plentiful or when much of the housing stock is rented. In contrast, prices aggregate beliefs well in cities where the housing stock is fixed and owner-occupied. In these areas, house prices reflect the average of all resident beliefs, even though they are agreeing to disagree and short-selling housing is impossible.

The paper proceeds as follows. In Section 2.2, we document the puzzling aspects of the cross-section of the U.S. housing boom, as well as the importance of supply-side speculation in land markets. Section 2.3 models the housing market environment. Section 2.4 contains our analysis of how house prices aggregate beliefs. In Section 2.5, we derive implications of the model to explain the empirical cross-section of housing markets during the U.S. boom. Section 2.6 contains new predictions on the cross-section of neighborhoods within a city, and Section 2.7 concludes.

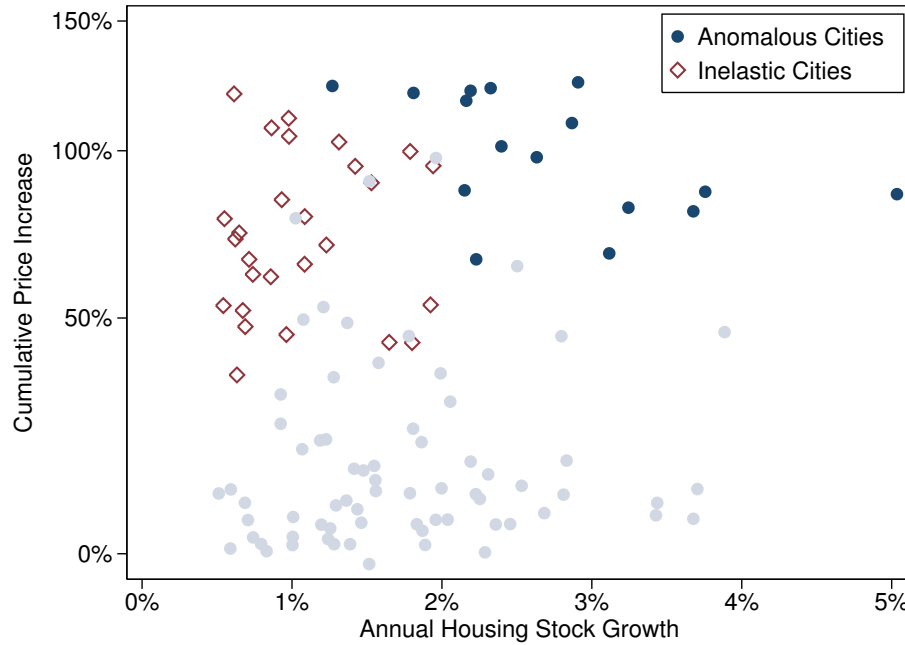
2.2 Stylized Facts of the U.S. Housing Boom and Bust

2.2.1 The Cross-Section of Cities

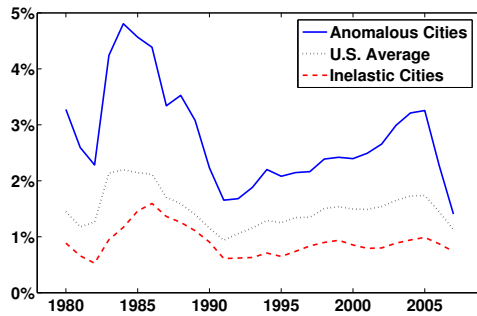
The Introduction mentions three puzzles about the cross-section of city experiences during the boom. First, large house price booms occurred in elastic cities where new construction historically had kept prices low. Second, the price booms in these elastic areas were as large as, if not larger than, those happening in inelastic cities at the same time. Finally, house prices remained flat in other elastic cities that were also rapidly building housing.

We document these puzzles using city-level house price and construction data. House price data come from the Federal Housing Finance Agency's metropolitan statistical area quarterly house price indices. We measure the housing stock in each city at an annual frequency by interpolating the U.S. Census's decadal housing stock estimates with its annual housing permit figures. Throughout, we focus on the 115 metropolitan areas for which the population in 2000 exceeds 500,000. The boom consists of the period between 2000 and 2006, matching the convention in the literature to use 2006 as the end point (Mian, Rao and Sufi, 2013).

a) Price Increases and Construction, 2000-2006



b) Historic Construction



c) Historic Prices

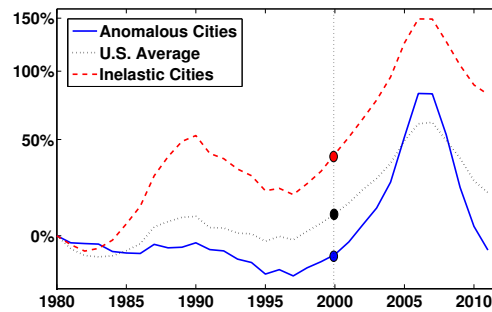


Figure 2.2: The U.S. Housing Boom and Bust Across Cities

Notes: Anomalous Cities include those in Arizona, Nevada, Florida, and inland California. Inelastic Cities are Boston, Providence, New York, Philadelphia, and all cities on the west coast of the United States. We measure the housing stock in each city at an annual frequency by interpolating the U.S. Census's decadal housing stock estimates with its annual housing permit figures. House price data come from the second quarter FHFA house price index deflated by the CPI-U. The figure includes all metropolitan areas with populations over 500,000 in 2000 for which we have data. (a) The cumulative price increase is the ratio of the house price in 2006 to the house price in 2000. The annual housing stock growth is the log difference in the housing stock in 2006 and 2000 divided by six. (b), (c) Each series is an average over cities in a group weighted by the city's housing stock in 2000. Construction is annual permitting as a fraction of the housing stock. Prices represent the cumulative returns from 1980 on the housing in each group.

Figure 2.2(a) plots construction and house price increases across cities during the boom. The house price increases vary enormously across cities, ranging from 0% to 125% over this brief six-year period. The largest price increases occurred in two groups of cities. The first group, which we call the Anomalous Cities, consists of Arizona, Nevada, Florida, and inland California. The other large price booms happened in the Inelastic Cities, which comprise Boston, Providence, New York, Philadelphia, and the west coast of the United States.

The history of construction and house prices in the Anomalous Cities before 2000 constitute the first puzzle. As shown in Figures 2.2(b) and 2.2(c), from 1980 to 2000 these cities provided clear examples of elastic housing markets in which prices stay low through rapid construction activity. Construction far outpaced the U.S. average while house prices remained constant. The standard model of housing cycles would have predicted the surge in U.S. housing demand between 2000 and 2006 to increase construction in these cities but not to raise prices. Empirically, the shock did increase construction, as shown in Panel (b). The puzzle is that house prices rapidly increased as well.

The second puzzle is that the price increases in the Anomalous Cities were as large as those in the Inelastic Cities. The Inelastic Cities consist of markets where house prices rise because regulation prohibits construction from absorbing higher demand. We document this relationship in Panels (b) and (c) of Figure 2.2, which show that construction in these cities was lower than the U.S. average before 2000 while house price growth greatly exceeded the U.S. average. The standard housing cycle model would have predicted the Inelastic Cities to lead the nation in house price growth in the boom after 2000. Although house prices did sharply rise, the price increases in the Inelastic Cities were no larger than those in the Anomalous Cities where the boom led to rapid construction.

The final puzzle is that some elastic cities built housing quickly during the boom but, unlike the Anomalous Cities, experienced stable house prices. These cities appear in the bottom-right corner of Figure 2.2(a), and are located in the southeastern United States (e.g. Texas and North Carolina). Their construction during the boom quantitatively matches that in the Anomalous Cities, but the price changes are significantly smaller. Why was rapid

construction able to hold down house prices in some cities and not others?

One response to these three puzzles is that the Anomalous Cities simply experienced much larger demand shocks than the rest of the nation during the boom. Although differential demand shocks surely explain part of the cross-section, they cannot account for all aspects of the Anomalous Cities just documented. These cities had been experiencing abnormally large demand shocks for years before 2000. Figure 2.2(b) shows that they were some of the fastest growing cities in the United States. Yet the surging demand to live in these areas did not increase prices. The departure from this pattern after 2000 requires a more nuanced theory than the hypothesis that housing demand increased particularly strongly in the Anomalous cities during the boom.

2.2.2 The Central Importance of Land Prices

This paper argues that speculation in land markets explains the variation in the house price boom across cities just documented. Our model demonstrates that land market speculation amplifies house price increases by making prices look more optimistic, and that this amplification is strongest in areas at the same level of development as the Anomalous Cities. In our framework, all movements in house prices arise from changes in land prices that reflect optimistic beliefs. Matching this premise, land price increases empirically account for nearly all of the increase in house prices during the boom, as we now show.

Tracing house price increases to land prices distinguishes our argument from “time-to-build” theories. According to the time-to-build hypothesis, house prices rise during a boom because of a temporary failure of homebuilders to expand construction. This delivery lag derives from obstacles erected by local regulators or from temporary shortages of inputs such as drywall and skilled labor. Under this theory, the price of undeveloped land should remain constant during the boom. Because land prices reflect the long-run, temporary housing shortages have no effect on the price of undeveloped land. These shortages instead raise construction costs and the shadow price of regulatory building permission.

To assess the importance of land prices, we gather data on land prices and construction

costs at the city level. Data on land prices come from the indices developed by Nichols, Oliner and Mulhall (2013) using land parcel transaction data. They run hedonic regressions to control for parcel characteristics and then derive city-level indices from the coefficients on city-specific time dummies. We measure construction costs using the R.S. Means construction cost survey. This survey asks homebuilders in each city to report the marginal cost of building a square foot of housing, including all labor and materials costs. Survey responses reflect real differences across cities in construction costs. In 2000, the lowest cost is \$54 per square foot and the highest is \$95; the mean is \$67 per square foot and the standard deviation is \$9.

Competition among homebuilders implies that, when construction is positive, house prices must equal land prices plus construction costs: $p_t^h = p_t^l + K_t$. Log-differencing this equation between 2000 and 2006 yields

$$\Delta \log p^h = \alpha \Delta \log p^l + (1 - \alpha) \Delta \log K,$$

where Δ denotes the difference between 2000 and 2006 and α is land's share of house prices in 2000. The factor that matters more should vary more closely with house prices across cities. Because α and $1 - \alpha$ are less than 1, the critical factor should also rise more than house prices do.

Figure 2.3 plots for each city the real growth in construction costs and land prices between 2000 and 2006 against the corresponding growth in house prices. Construction costs rose relatively little during this period, and growth in these costs does not vary in relation to the size of house price increases. Land prices display the opposite pattern, rising substantially during the boom and exhibiting a high correlation with house prices. Each city's land price increase also exceeds its house price increase. This evidence underscores the central importance of land prices for understanding the cross-section of house price booms.

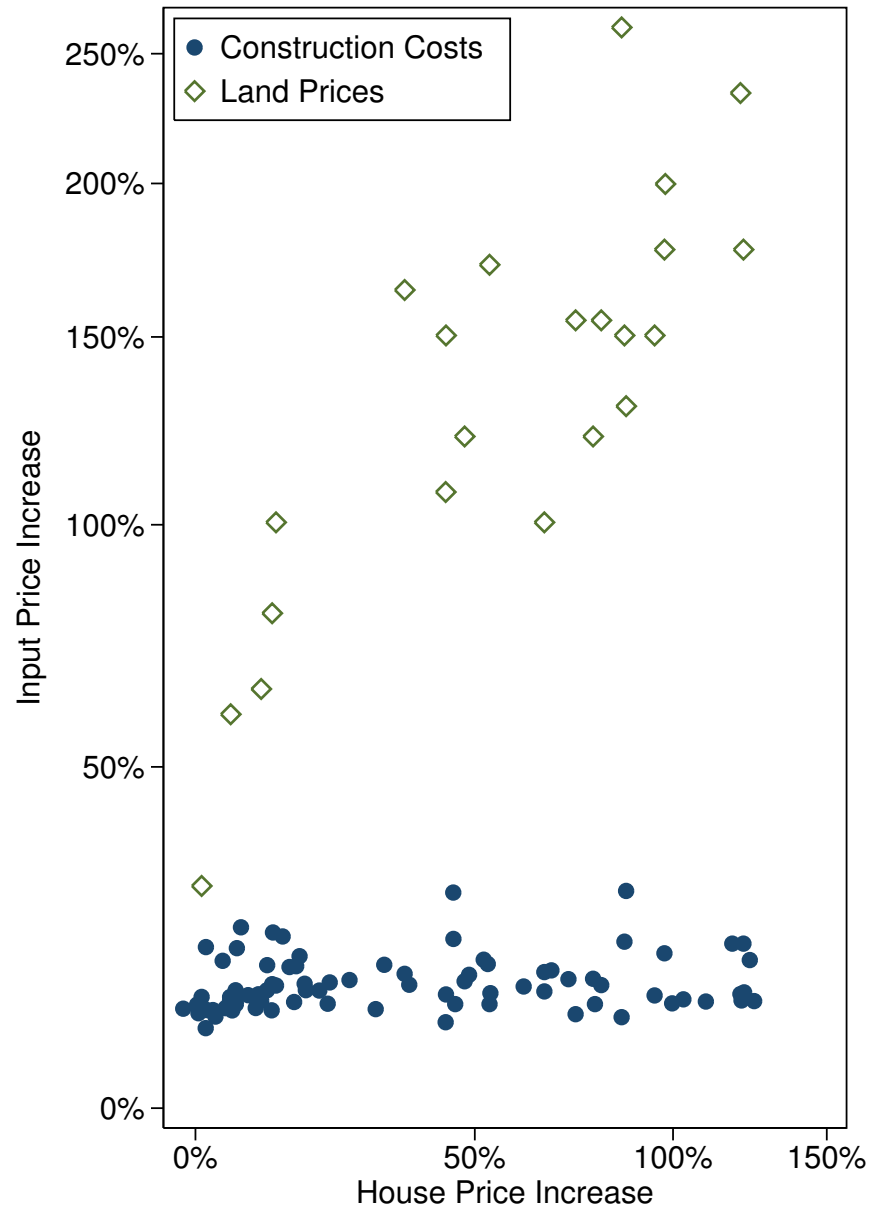


Figure 2.3: *Input Price and House Price Increases Across Cities, 2000-2006*

Notes: We measure construction costs for each city using the R.S. Means survey figures for the marginal cost of a square foot of an average quality home, deflated by the CPI-U. Gyourko and Saiz (2006) contains further information on the survey. Land price changes come from the hedonic indices calculated in Nichols, Oliner and Mulhall (2013) using land parcel transactions, and house prices come from the second quarter FHFA housing price index deflated by the CPI-U. The figure includes all metropolitan areas with populations over 500,000 in 2000 for which we have data.

2.2.3 Land Market Speculation by Homebuilders

The land price booms just documented were driven by speculation in land markets. The term “speculation” refers to the process in which optimists buy up an asset that cannot be shorted, biasing its price. Our model describes two implications of this behavior. First, the owners of the land during the boom increase their positions as they crowd out less optimistic landowners. Second, when their beliefs are revealed to be more optimistic than reality, optimists suffer capital losses. We document both of these features among a class of landowners for whom rich data are publicly available: public homebuilders. We focus on the eight largest firms and hand-collect landholding data from their annual financial statements between 2001 and 2010.

Consistent with speculative behavior, these firms nearly tripled their landholdings between 2001 and 2005, as shown in Figure 2.4(a). These land acquisitions far exceed additional land needed for new construction. Annual home sales increased by 120,000 between 2001 and 2005, while landholdings increased by 1,100,000 lots. One lot can produce one house, so landholdings rose more than *nine times* relative to home sales. In 2005, Pulte changed the description of its business in its 10-K to say, “We consider land acquisition one of our core competencies.” This language appeared until 2008, when it was replaced by, “Homebuilding operations represent our core business.”

Having amassed large land portfolios, these firms subsequently suffered large capital losses. Figure 2.4(b) documents the dramatic rise and fall in the total market equity of these homebuilders between 2001 and 2010. Homebuilder stocks rose 430% and then fell 74% over this period. The majority of the losses borne by homebuilders arose from losses on the land portfolios they accumulated from 2001 to 2005. In 2006, these firms began reporting write-downs to their land portfolios. At \$29 billion, the value of the land losses between 2006 and 2010 accounts for 73% of the market equity losses over this time period. The homebuilders bore the entirety of their land portfolio losses. The absence of a hedge against downside risk supports the theory that homebuilder land acquisitions represented their optimistic beliefs.



Figure 2.4: Supply-Side Speculation Among U.S. Public Homebuilders, 2001-2010

Notes: (a), (b) Data come from the 10-K filings of Centex, Pulte, Lennar, D.R. Horton, K.B. Homes, Toll Brothers, Hovnanian, and Southern Pacific, the eight largest public U.S. homebuilders in 2001. "Lots Controlled" equals the sum of lots directly owned and those controlled by option contracts. The cumulative writedowns to land holdings between 2006 and 2010 among these homebuilders totals \$29 billion. (c) Short interest is computed as the ratio of shares currently sold short to total shares outstanding. Monthly data series for shares short come from COMPUSTAT and for shares outstanding come from CRSP. Builder stocks are classified as those with NAICS code 236117.

Further evidence of homebuilder optimism comes from short-selling of their market equity. If the homebuilders buying land are more optimistic than most investors, then other investors should bet against them by shorting their stock. Figure 2.4(c) plots monthly short interest ratios, defined as the ratio of shares currently sold short to total shares outstanding, for homebuilder stocks and non-homebuilder stocks between 2001 and 2010. Throughout the boom, short interest of homebuilder stock sharply increased, rising from 2% in 2001 to 12% in 2006. It further increased as homebuilders began to announce their land losses in 2006. Rising short interest provides direct evidence of disagreement over the value of homebuilder land portfolios and thus over future house prices.

2.3 A Housing Market with Homeowners and Developers

Housing Supply. The city we study has a fixed amount of space S . This space can either be used for housing, or it remains as undeveloped land. The total housing stock in the city at time t is H_t and the remaining undeveloped land is L_t , so $S = H_t + L_t$ for all t .

A continuum of real estate developers invest in land and construct housing from the land at a cost of K per unit of housing. The aggregate supply of new housing is ΔH_t . Construction is instantaneous, and housing does not depreciate: $H_t = \Delta H_t + H_{t-1}$. Construction is also irreversible: $\Delta H_t \geq 0$. Both housing and land are continuous variables, and one unit of housing requires one unit of land.

The developers rent out land on spot markets at a price of r_t^l . Rental demand for undeveloped land comes from firms, such as farms, that use the city's land as an input. These firms buy their inputs and sell their products on the global market. Therefore, their aggregate demand for land depends only on r_t^l and not on any other local market conditions. This aggregate rental demand curve is $D^l(r_t^l)$, where $D^l(\cdot)$ is decreasing positive function such that $D^l(0) \geq S$.

The profit flow of a developer j at time t is

$$\pi_{j,t} = \underbrace{r_t^l L_{j,t} + p_t^l (L_{j,t-1} - L_{j,t})}_{\text{development profit}} + \underbrace{(p_t^h - p_t^l - K) \Delta H_{j,t}}_{\text{homebuilding profit}}, \quad (2.1)$$

where p_t^h is the price of housing and p_t^l is the price of land. The real estate development industry faces no entry costs, so the industry is perfectly competitive. Because homebuilding is instantaneous and does not depend on prior land investments, profits from this line of business must be zero due to perfect competition. We denote the aggregate homebuilding profit by $\pi_t^{hb} = (p_t^h - p_t^l - K) \Delta H_t$.

Each developer begins with a land endowment and issues equity to finance its land investments. It maximizes its expected net present value of profits $E_j \sum_{t=0}^{\infty} \beta^t \pi_{j,t}$. The operator E_j reflects firm j 's expectation of future land prices. Firm-specific beliefs represent the beliefs of the firm's CEO, who owns equity, cannot be fired, and decides the firm's land investments. The number of each developer's equity shares equals the amount of land it holds, and each developer pays out its land rents as dividends. The market price of developer equity therefore equals the market price p_t^l of land.

Individual Housing Demand. A population of residents live in the city and hold its housing. These residents receive direct utility from consuming housing. Lower-case h denotes the flow consumption of housing, whereas upper-case H denotes the asset holding. Flow utility from housing depends on whether housing is consumed through owner-occupancy or under a rental contract. Residents also derive utility from non-local consumption c . Each resident i maximizes the expected present value of utility, given by

$$E_i \sum_{t=0}^{\infty} \beta^t u_i(c_t, h_t^{own}, h_t^{rent}),$$

where β is the common discount factor.

Flow utility $u_i(\cdot, \cdot, \cdot)$ has three properties. First, it is separable and linear in non-real estate consumption c . This quasi-linearity eliminates risk aversion and hedging motives. Second, owner-occupied and rented housing are substitutes, and residents vary in which

type of contract they prefer and to what degree. Substitutability of owner-occupied and rented housing fully sorts residents between the two types of contracts; no resident consumes both types of housing simultaneously. Finally, residents face diminishing marginal utility of owner-occupied housing. This property leads homeownership to be dispersed among residents in equilibrium.

The utility specification we adopt that features these three properties is

$$u_i(c, h^{own}, h^{rent}) = c + v(a_i h^{own} + h^{rent}) \quad (2.2)$$

where $a_i > 0$ is resident i 's preference for owner-occupancy, and $v(\cdot)$ is an increasing, concave function for which $\lim_{h \rightarrow 0} v'(h) = \infty$. The distribution of the owner-occupancy preference parameter a_i across residents is given by a continuously differentiable cumulative distribution function F_a , which is stable over time. Owner-occupancy utility is unbounded: dF_a has full support on \mathbb{R}^+ . The functional form of the owner-occupancy preference in (2.2) results from a moral hazard problem we describe in the Appendix.

Resident Optimization. Residents hold three assets classes: bonds B , housing H , and developer equity Q . Global capital markets external to the city determine the gross interest rate on bonds, which is $R_t = 1/\beta$, where β is the common discount factor. Residents may borrow or lend at this rate by buying or selling these bonds in unlimited quantities.

In contrast, housing and developer equity are traded within the city, and equilibrium conditions determine their prices p_t^l and p_t^h . Homeowners earn income by renting out the housing they own in excess of what they consume. The spot rental price for housing is r_t^h ; landlord revenue is therefore $r_t^h(H_{i,t} - h_{i,t}^{own} - h_{i,t}^{rent})$. Shorting housing is impossible, but residents can short developer equity. Doing so is costly. Residents incur a convex cost $k_s(Q)$ to short Q units of developer stock, where $k_s(0) = 0$ and $k'_s, k''_s > 0$. These costs reflect fees paid to borrow stock, as well as time spent locating available stock (D'Avolio, 2002).

Short-sale constraints in the housing market result from a lack of asset interchangeability. Although housing is homogeneous in the model, empirical housing markets involve large variation in characteristics across houses. This variation in characteristics makes it essentially

impossible to cover a short. Unlike in the housing market, asset interchangeability holds in the equity market, where all of a firm's shares are equivalent.

The Bellman equation representing the resident optimization problem is

$$V(B_{i,t-1}, H_{i,t-1}, Q_{i,t-1}) = \max_{\substack{B_{i,t}, H_{i,t}, Q_{i,t} \\ c_{i,t}, h_{i,t}^{own}, h_{i,t}^{rent}}} \underbrace{c_{i,t} + v(a_i h_{i,t}^{own} + h_{i,t}^{rent})}_{\text{flow utility}} + \underbrace{\beta \mathbf{E}_{i,t} V(B_{i,t}, H_{i,t}, Q_{i,t})}_{\text{continuation value}}, \quad (2.3)$$

where the maximization is subject to the short-sale constraint

$$0 \leq H_{i,t},$$

the ownership constraint

$$h_{i,t}^{own} \leq H_{i,t},$$

and the budget constraint

$$\underbrace{R_t B_{i,t-1} - B_{i,t}}_{\text{borrowing costs}} + \underbrace{c_{i,t}}_{\text{consumption}} \leq \underbrace{p_t^h (H_{i,t-1} - H_{i,t})}_{\text{housing returns}} + \underbrace{r_t^h (H_{i,t} - h_{i,t}^{own} - h_{i,t}^{rent})}_{\text{housing rental income}} \\ + \underbrace{p_t^l (Q_{i,t-1} - Q_{i,t})}_{\text{equity returns}} + \underbrace{r_t^l Q_{i,t}}_{\text{dividends}} - \underbrace{\max(0, k_s(-Q_{i,t}))}_{\text{shorting costs}}.$$

Aggregate Demand and Beliefs. Aggregate demand to live in the city equals the number of residents N_t . This aggregate demand consists of a shock and a trend:

$$\underbrace{\log N_t}_{\text{demand}} = \underbrace{z_t}_{\text{shock}} + \underbrace{\log \bar{N}_t}_{\text{trend}}.$$

The trend component grows at a constant positive rate g : for all $t > 0$,

$$\log \bar{N}_t = g + \log \bar{N}_{t-1}.$$

The shocks z_t have a common factor x . The dependence of the time- t shock on the common factor x is μ_t , so that

$$z_t = \mu_t x.$$

Without loss of generality, $\mu_0 = 1$: the time 0 shock z_0 equals the common factor x . We denote $\boldsymbol{\mu} = \{\mu_t\}_{t \geq 0}$.

At time 0, residents observe the following information: the current and future values of trend demand \bar{N}_t , the trend growth rate g , the current demand N_0 , the current shock z_0 , and the common factor x of the future shocks. They do *not* observe $\boldsymbol{\mu}$, the data needed to extrapolate the factor x to future shocks. Residents learn the true value of the entire vector $\boldsymbol{\mu}$ at time $t = 1$. The resolution of uncertainty at time $t = 1$ is common knowledge at $t = 0$.

Residents agree to disagree about the true value of $\boldsymbol{\mu}$. At time 0, resident i 's subjective prior of $\boldsymbol{\mu}$ is given by F_i , an integrable probability measure on the compact space M of all possible values of $\boldsymbol{\mu}$. These priors vary across residents. The resulting subjective expected value of each μ_t is $\mu_{i,t} = \int_M \mu_t dF_i$, and the vector of resident i 's subjective expected values of each μ_t is $\boldsymbol{\mu}_i = \{\mu_{i,t}\}_{t \geq 0}$. The subjective expected value $\boldsymbol{\mu}_i$ uniquely determines the prior F_i . The distribution of $\boldsymbol{\mu}_i$ itself across residents admits an integrable probability distribution F_μ on M , which is independent from the distribution F_a of owner-occupancy preferences. The CEOs of the development firms are city residents, so their beliefs are drawn from the same distribution F_μ .

Resident disagreement reflects the unprecedented nature of the demand shock \mathbf{z} . As argued by Morris (1996), this heterogeneous prior assumption is most appropriate when investors face an unprecedented situation in which they have not yet had a chance to collect information and engage in rational updating. The events surrounding housing booms are precisely these types of situations. Glaeser (2013) meticulously shows that in each of the historical booms he analyzes, reasonable investors could agree to disagree about future real estate prices. In the case of the U.S. housing boom between 2000 and 2006, we follow Mian and Sufi (2009) in thinking of the shock as the arrival of new securitization technologies that expanded credit to low-income borrowers. The initial shock to housing demand is x , and $\boldsymbol{\mu}$ represents the degree to which this expansion of credit in 2000-2006 persists after 2006.

Equilibrium. Equilibrium consists of time-series vectors of prices $\mathbf{p}^L(\boldsymbol{\mu})$, $\mathbf{p}^H(\boldsymbol{\mu})$, $\mathbf{r}^l(\boldsymbol{\mu})$, $\mathbf{r}^h(\boldsymbol{\mu})$ and quantities $\mathbf{L}(\boldsymbol{\mu})$, $\mathbf{H}(\boldsymbol{\mu})$ that depend on the realized value of $\boldsymbol{\mu}$. These pricing and

quantity functions constitute an equilibrium when housing, land, and equity markets clear while residents and developers maximize utility and profits:

Consider pricing functions $\mathbf{p}^h(\boldsymbol{\mu}), \mathbf{p}^l(\boldsymbol{\mu}), \mathbf{r}^h(\boldsymbol{\mu}), \mathbf{r}^l(\boldsymbol{\mu})$ and quantity functions $\mathbf{H}(\boldsymbol{\mu}), \mathbf{L}(\boldsymbol{\mu})$. Let $H_{i,t}^*, Q_{i,t}^*, (h_{i,t}^{own})^*$, and $(h_{i,t}^{rent})^*$ be resident i 's solutions to the Bellman equation (2.3) given his owner-occupancy preference a_i , his beliefs $\boldsymbol{\mu}_i$, and these pricing functions. Let $L_{j,t}^*$ be developer j 's land holdings that maximize expected net present value of profits in equation (2.1), given the pricing functions; L_t^* is the sum of these land holdings across developers. The pricing and quantity functions constitute a recursive competitive equilibrium if at each time t :

1. The sum of undeveloped land and housing equals the city's endowment of open space:

$$S = L_t(\boldsymbol{\mu}) + H_t(\boldsymbol{\mu}).$$

2. Flow demand for land equals investment demand from developers, which equals the resident demand for their equity:

$$L_t(\boldsymbol{\mu}) = L_t^* = D^l(r_t^l(\boldsymbol{\mu})) = \int_0^\infty \int_M Q_{i,t}^* dF_\mu dF_a.$$

3. Resident stock and flow demand for housing clear:

$$H_t(\boldsymbol{\mu}) = N_t(\boldsymbol{\mu}) \int_0^\infty \int_M H_{i,t}^* dF_\mu dF_a = N_t(\boldsymbol{\mu}) \int_0^\infty \int_M ((h_{i,t}^{own})^* + (h_{i,t}^{rent})^*) dF_\mu dF_a.$$

4. Construction maximizes developer profits:

$$H_t(\boldsymbol{\mu}) - H_{t-1}(\boldsymbol{\mu}) \in \arg \max_{\Delta H_t} \pi_t^{hb}.$$

5. Developer profit from homebuilding is zero:

$$\max_{\Delta H_t} \pi_t^{hb} = 0.$$

Elasticity of Housing Supply. The housing supply curve is the city's open space S less the rental demand for land $D^l(r_t^l)$. We denote the elasticity of this supply curve with respect to housing rents r_t^h by ϵ_t^S . The supply elasticity determines the construction response to the

shocks $\{z_t\}$. It will also serve as a sufficient statistic for the extent to which land speculation affects house prices. This section describes the supply elasticity ϵ_t^S along the city's trend growth path, which obtains when $x = 0$.

The relationship between land rents r_t^l and house rents r_t^h allows us to calculate this elasticity. Because trend growth $g > 0$, new residents perpetually arrive to the city, and developers build new houses each period. Perpetual construction ties together land and house prices. In particular, as developers must be indifferent between building today or tomorrow, house rents equal land rents plus flow construction costs:

$$r_t^h = r_t^l + (1 - \beta)K.$$

The supply of housing is open space net of flow land demand: $S - D^l(r_t^h - (1 - \beta)K)$. The elasticity of housing supply is thus $\epsilon_t^S \equiv -r_t^h(D^l)' / (S - D^l)$. When the flow land demand D^l features a constant elasticity ϵ^l , the elasticity of housing supply takes on the simple form

$$\epsilon_t^S = \frac{r_t^h}{r_t^h - (1 - \beta)K} \left(\frac{S}{H_t} - 1 \right) \epsilon^l, \quad (2.4)$$

where H_t is the housing stock at time t . The arrival of new residents increases both rents r_t^h and the level of development H_t/S . The supply elasticity given in (2.4) unambiguously falls (see Appendix for proof):

Lemma 1 *Define housing supply to be the residual of the city's open space S minus the flow demand for land: $S - D^l$. The elasticity ϵ_t^S of housing supply with respect to housing rents r_t^h decreases with the level of city development H_t/S along the city's trend growth path.*

2.4 Supply-Side Speculation

At time 0, residents disagree about the future path of housing demand. Speculative trading behavior results from this disagreement. This section describes how owner-occupancy frictions crowd speculators out of owner-occupied housing and into rental housing and land. Demand and supply elasticities determine how prices aggregate the beliefs of owner-

occupants and of optimistic speculators holding rental housing and land.

2.4.1 Land Speculation and Dispersed Homeownership

We first consider the developer decision to hold land at time 0. Developer j 's first-order condition on its land-holding $L_{j,0}$ is

$$\underbrace{1/\beta}_{\text{risk-free rate}} \geq \underbrace{\mathbf{E}_j p_1^l / (p_0^l - r_0^l)}_{\text{expected land return}},$$

with equality if and only if $L_{j,0} > 0$. A developer invests in land if and only if it expects land to return the risk-free rate. At time 0, developers disagree about this expected return on land because they disagree about the future path of housing demand. The developers that expect the highest returns invest in land, while all other developers sell to these optimistic firms and exit the market. We denote the optimistic belief of the developers who invest in land by $\tilde{\mathbf{E}} p_1^l \equiv \max_{\mu_j} \mathbf{E}(p_1^l \mid \mu_j)$.

Optimistic residents finance developer investments in land through purchasing their equity. Less optimistic residents choose to short-sell developer stock. Developer stock allows residents to hold land indirectly: its price is p_0^l and it pays a dividend of r_0^l . Resident i holds this equity only if he agrees with the land valuation of the optimistic developers, in which case $\mathbf{E}_i p_1^l = \tilde{\mathbf{E}} p_1^l$. Otherwise, he shorts the equity, and his first-order condition is

$$k'_s(-Q_{i,0}^*) = \beta(\tilde{\mathbf{E}} p_1^l - \mathbf{E}_i p_1^l).$$

Disagreement increases the short interest in the equity of the developers holding the land. Without disagreement, $\tilde{\mathbf{E}} p_1^l = \mathbf{E}_i p_1^l$ for all residents, so no one shorts.

Only the most optimistic residents hold housing as landlords. A resident is a *landlord* if he owns more housing than he consumes through owner-occupancy: $H_i > h_i^{own}$. The first-order condition of the Bellman equation (2.3) with respect to $H_{i,0}$ when it is in excess of $h_{i,0}^{own}$ is

$$\underbrace{1/\beta}_{\text{risk-free rate}} \geq \underbrace{\mathbf{E}_i p_1^h / (p_0^h - r_0^h)}_{\text{expected housing return}}, \quad (2.5)$$

with equality if and only if $H_{i,0} > h_{i,0}^{own}$. Only the most optimistic residents invest in rental housing, just as only the most optimistic developers invest in land. Land and rental housing share this fundamental property. During periods of uncertainty, the most optimistic investors are the sole holders of these asset classes.

Owner-occupancy utility crowds these optimistic investors out of owner-occupied housing, which remains dispersed among residents of all beliefs. The decision to own or rent emerges from the first-order conditions of the Bellman equation (2.3) with respect to $h_{i,0}^{own}$ and $h_{i,0}^{rent}$. We express these equations jointly as

$$\underbrace{v'(a_i(h_{i,0}^{own})^* + (h_{i,0}^{rent})^*)}_{\text{marginal utility of housing}} = \min \left(\underbrace{a_i^{-1}(p_0^h - \beta \mathbf{E}_i p_1^h)}_{\text{owning}}, \underbrace{r_0^h}_{\text{renting}} \right). \quad (2.6)$$

The left term in the parentheses denotes the expected flow price of marginal utility v' from owning a house; the right term denotes the flow price of renting. A resident owns when the owner-occupancy price is less than the rental price. As long as the owner-occupancy preference a_i is large enough, resident i decides to own even if his belief $\mathbf{E}_i p_1^h$ is quite pessimistic. Homeownership remains dispersed among residents of all beliefs.

We gain additional intuition about the own-rent margin by substituting (2.5) into (2.6). We denote the most optimistic belief about future house prices, the one held by landlords investing in rental housing, by $\tilde{\mathbf{E}} p_1^h \equiv \max_{\mu_i} \mathbf{E}(p_1^h \mid \mu_i)$. The decision to own rather than rent reduces to

$$a_i \geq 1 + \frac{\beta(\tilde{\mathbf{E}} p_1^h - \mathbf{E}_i p_1^h)}{r_0^h}. \quad (2.7)$$

Without disagreement, a resident owns exactly when he intrinsically prefers owning to renting, so that $a_i \geq 1$. Disagreement sets the bar higher. Some pessimists for whom $a_i \geq 1$ choose to rent because they expect capital losses on owning a home. Other pessimists continue to own because their owner-occupancy utility is high enough to offset the fear of capital losses. Proposition 1 summarizes these results.

Proposition 1 *Owner-occupancy utility crowds speculators out of the owner-occupied housing market and into the land and rental markets. The most optimistic residents—those holding the highest*

value of $E_i p_1^h$ —buy up all rental housing and finance optimistic developers who purchase all the land. In contrast, owner-occupied housing remains dispersed among residents of all beliefs.

Proposition 1 yields two corollaries that match stylized facts presented in Section 2.2. The most optimistic developers buy up all the land. Unless they start out owning all the land, these optimistic developers increase their land positions following the demand shock. They hold this land as an investment rather than for immediate construction.

Implication 1 *The developers who hold land at time 0 increase their aggregate land holdings at time 0. They buy land in excess of their immediate construction needs.*

This implication explains the land-buying activities of large public U.S. homebuilders documented in Figure 2.4(a).

The second corollary concerns short-selling. Residents who disagree with the optimistic valuations of developers short their equity.

Implication 2 *Disagreement increases the short interest of developer equity at time 0.*

Figure 2.4(c) documents the rising short interest in the stocks of U.S. public homebuilders who were taking on large land positions during the boom. This short interest provides direct evidence of disagreement during the boom.

2.4.2 Belief Aggregation

Prices aggregate the heterogeneous beliefs of residents and developers holding housing and land. The real estate market consists of three components: land, rental housing, and owner-occupied housing. The most optimistic residents hold the first two, while the third remains dispersed among owner-occupants. House prices reflect a weighted average of the optimistic belief and the average belief of all owner-occupants. The weight on the optimistic belief is the share of the real estate market consisting of land and rental housing; the weight on the average owner-occupant belief is owner-occupied housing's share of the market.

To derive these results, we take a comparative static of the form $\partial p_0^h / \partial x$. The shock $\mathbf{z} = \mu x$ scales with the common factor x . We differentiate with respect to x at $x = 0$ to

explore how prices change as the shocks, and hence the ensuing disagreement, increase. Our partial derivative holds current demand N_0 constant to isolate the aggregation of future beliefs.

We first use (2.5) to write $p_0^h = r_0^h + \beta \tilde{\mathbf{E}} p_1^h$. The shock increases the optimistic belief $\beta \tilde{\mathbf{E}} p_1^h$, directly increasing prices. It also changes the market rent r_0^h . This rent is determined by the intersection of housing supply and housing demand:

$$\underbrace{S - D^l \left(r_0^h - (1 - \beta)K \right)}_{\text{housing supply}} = \underbrace{D_0^h(r_0^h)}_{\text{housing demand}}, \quad (2.8)$$

where

$$\begin{aligned} D_0^h(r_0^h) = & \underbrace{N_0 \int_M \int_0^{1+\beta(\tilde{\mathbf{E}} p_1^h - \mathbf{E}_i p_1^h)/r_0^h} (v')^{-1}(r_0^h) dF_a dF_\mu}_{\text{rental housing}} \\ & + \underbrace{N_0 \int_M \int_{1+\beta(\tilde{\mathbf{E}} p_1^h - \mathbf{E}_i p_1^h)/r_0^h}^{\infty} a_i^{-1}(v')^{-1} \left(a_i^{-1}(r_0^h + \beta(\tilde{\mathbf{E}} p_1^h - \mathbf{E}_i p_1^h)) \right) dF_a dF_\mu}_{\text{owner-occupied housing}}. \end{aligned} \quad (2.9)$$

The housing demand equation follows from (2.6) and (2.7). We determine the shock's effect on rents by totally differentiating (2.8) with respect to x at $x = 0$, keeping current demand N_0 constant. When the elasticity of housing demand ϵ^D is constant, the resulting comparative static $\partial p_0^h / \partial x$ adopts the simple form given in the following proposition, which we prove in the Appendix.

Proposition 2 *Consider the partial effect of the shock in which current demand N_0 stays constant but future house price expectations $\mathbf{E}_i p_1^h$ change. The change in house prices averages the changes in the optimistic resident belief and the average belief:*

$$\frac{\partial p_0^h}{\partial x} = \frac{\epsilon_0^S + (1 - \chi)\epsilon^D}{\epsilon_0^S + \epsilon^D} \frac{\partial \beta \tilde{\mathbf{E}} p_1^h}{\partial x} + \frac{\chi \epsilon^D}{\epsilon_0^S + \epsilon^D} \frac{\partial \beta \bar{\mathbf{E}} p_1^h}{\partial x}, \quad (2.10)$$

where $\tilde{\mathbf{E}} p_1^h = \max_i \mathbf{E}_i p_1^h$ is the most optimistic belief, $\bar{\mathbf{E}} p_1^h = \int_M \mathbf{E}_i p_1^h dF_\mu$ is the average belief, ϵ_0^S is the elasticity of housing supply at time 0, ϵ^D is the elasticity of housing demand, and $\chi = \int_0^\infty (h_{i,0}^{\text{own}})^* dF_a / H_0$ is the share of housing that is owner-occupied when $x = 0$.

The weight on the optimistic belief in Proposition 2 represents the share, on the margin, of the real estate market owned by speculators. The supply elasticity ϵ_0^S represents land, and $(1 - \chi)\epsilon^D$ represents rental housing. The remaining $\chi\epsilon^D$ represents owner-occupied housing and is the weight on the average owner-occupant belief. The average owner-occupant belief coincides with the unconditional average belief because at $x = 0$, beliefs and tenure choice are independent.

Proposition 2 yields four corollaries on the difference in belief aggregation across cities and neighborhoods. Prices look more optimistic when the weight $(\epsilon_0^S + (1 - \chi)\epsilon^D) / (\epsilon_0^S + \epsilon^D)$ is higher. This ratio is greater when the supply elasticity ϵ_0^S is higher:

Implication 3 *Prices look more optimistic when the housing supply elasticity is higher, i.e. in less developed cities.*

Disagreement reverses the common intuition relating housing supply elasticity and movements in house prices. Elastic supply keeps prices low by allowing construction to respond to demand shocks. But land constitutes a larger share of the real estate market when supply is elastic. Speculators are drawn to the land markets, so elastic supply amplifies the role of speculators in determining prices during periods of disagreement. When supply is perfectly elastic, $\epsilon_0^S = \infty$ and prices reflect only the beliefs of these optimistic speculators:

Implication 4 *When housing supply is perfectly elastic, house prices incorporate only the most optimistic beliefs; they reflect the beliefs of developers and not of owner-occupants.*

Recent research has measured owner-occupant beliefs about the future evolution of house prices.⁸ In cities with elastic housing supply, such as the cities motivating this paper, developer rather than owner-occupant beliefs determine prices. Data on the expectations of homebuilders would supplement the research on owner-occupant beliefs to explain prices in these elastic areas.

Prices aggregate beliefs much better when housing supply is perfectly inelastic ($\epsilon_0^S = 0$)

⁸See Landvoigt (2011), Case, Shiller and Thompson (2012), Burnside, Eichenbaum and Rebelo (2013), Soo (2013), Suher (2013), and Cheng, Raina and Xiong (2014).

and all housing is owner-occupied ($\chi = 1$). In this case, the price change depends only on the average belief $\bar{E}p_1^h$:

Implication 5 *When the housing stock is fixed and all housing is owner-occupied, prices reflect the average belief about long-run growth.*

In many settings, such as when investor information equals a signal plus mean zero noise, prices reflect all information when they incorporate the average private belief of all investors. Owner-occupied housing markets with a fixed housing stock display this property, even though short-selling is impossible and residents persistently disagree. These frictions fail to bias prices because homeownership remains dispersed among residents of all beliefs, due to the utility flows that residents derive from housing.

The weight $(\epsilon_0^S + (1 - \chi)\epsilon^D)/(\epsilon_0^S + \epsilon^D)$ on optimistic beliefs is also higher when χ is lower:

Implication 6 *Prices look more optimistic when a greater share of housing is rented.*

Speculators own a greater share of the real estate market when the rental share $1 - \chi$ is higher. Prices bias towards optimistic beliefs in market segments where more of the housing stock is rented.

2.5 The Cross-Section of City Experiences During the Boom

This section explains three puzzling aspects of the U.S. housing boom that occurred between 2000 and 2006. First, large house price booms occurred in elastic cities where new construction historically had kept prices low. Second, the price booms in these elastic areas were as large as, if not larger than, those happening in inelastic cities at the same time. Finally, house prices remained flat in other elastic cities that were also rapidly building housing.

To explain these cross-sectional facts, we derive a formula for the total effect of the shock \mathbf{z} on house prices. This formula expresses the house price boom as a function of the city's level of development when the shock occurs. Our analysis up to this point has

explored the partial effect of how prices aggregate beliefs $E_i p_1^h$, without specifying how these beliefs are formed. To derive the total effect of the shock, we express the changes in these beliefs in terms of city characteristics and the exogenous demand process. Specifically, we calculate the partial derivative $\partial \log p_0^h / \partial x$ holding all beliefs fixed at $\mu_i = \mu$, and then use Proposition 2 to derive the total effect of the shock x on house prices. As before, we evaluate derivatives at $x = 0$.⁹

At time 0, each resident expects the shock z_t to raise log-demand at time t by $\mu_t x$. The resulting expected change in rents r_t^h depends on the elasticities of supply and demand at time t :

$$\frac{\partial \log E_0 r_t^h}{\partial x} = \frac{\mu_t}{\epsilon_t^S + \epsilon^D}.$$

This equation follows from price theory. When a demand curve shifts up, a good's price increases by the inverse of the total elasticity of supply and demand. The total effects of the shocks $\{z_t\}$ on the current house price p_0^h follows from aggregating the above equation across all time periods, using the relation $p_0 = E_0 \sum_{t=0}^{\infty} \beta^t r_t^h$:

$$\frac{\partial \log p_0^h}{\partial x} = \frac{\mu}{\tilde{\epsilon}^S + \epsilon^D}. \quad (2.11)$$

The mean persistence of the shock is $\mu = \sum_{t=0}^{\infty} \mu_t \beta^t r_t^h (\epsilon_t^S + \epsilon^D)^{-1} / \sum_{t=0}^{\infty} \beta^t r_t^h (\epsilon_t^S + \epsilon^D)^{-1}$, and $\tilde{\epsilon}^S$ is the *long-run supply elasticity* given by the weighted harmonic mean of future supply elasticities in the city:

$$\tilde{\epsilon}^S \equiv -\epsilon^D + \frac{\sum_{t=0}^{\infty} \beta^t r_t^h}{\sum_{t=0}^{\infty} \beta^t r_t^h (\epsilon_t^S + \epsilon^D)^{-1}}.$$

The higher this long-run supply elasticity, the smaller the shock's impact on current house prices, holding μ fixed.

We now put together the two channels through which the shock changes prices. Equation (2.11) expresses the price change that results when μ is known, and (2.10) describes how prices aggregate residents' heterogeneous beliefs about μ . Proposition 3 states the total

⁹Evaluating derivatives at $x = 0$ describes the model when construction occurs in each period. When x is large enough and the shock \mathbf{z} might mean-revert, a construction stop at $t = 1$ is possible and anticipated by residents at $t = 0$. This feature of housing cycles distracts from our focus on housing booms and how they vary across cities.

effect $d \log p_0^h / dx$, which we formally calculate in the Appendix.

Proposition 3 *The total effect of the shock x on current house prices is*

$$\frac{d \log p_0^h}{dx} = \underbrace{\left(\frac{\epsilon_0^S + (1 - \chi)\epsilon^D}{\epsilon_0^S + \epsilon^D} \tilde{\mu} + \frac{\chi\epsilon^D}{\epsilon_0^S + \epsilon^D} \bar{\mu} \right)}_{\text{aggregate belief}} \underbrace{\frac{1}{\tilde{\epsilon}^S + \epsilon^D}}_{\text{pass-through}}, \quad (2.12)$$

where ϵ_0^S is the current elasticity of housing supply, $\tilde{\epsilon}^S$ is the long-run supply elasticity, ϵ^D is the elasticity of housing demand, χ is the share of housing that is owner-occupied, $\tilde{\mu}$ is the mean persistence of the most optimistic belief about μ , and $\bar{\mu}$ is the mean persistence of the average belief.

The first puzzle (2.12) explains is how a city with perfectly elastic housing supply can experience a house price boom. Housing supply is perfectly elastic when $\epsilon_0^S = \infty$. In this case, the house price boom is $\tilde{\mu}x / (\tilde{\epsilon}^S + \epsilon^D)$. This price increase is positive as long as the long-run supply elasticity $\tilde{\epsilon}^S$ is not also infinite.

Implication 7 *A house price boom occurs in a city where current housing supply is completely elastic, construction costs are constant, and construction is instantaneous. Supply must be inelastic in the future for such a price boom to occur.*

In the Appendix, we prove that a limiting case exists in which $\epsilon_0^S = \infty$ while $\tilde{\epsilon}^S < \infty$.

A house price boom results from a shock to current demand accompanied by news of future shocks. When supply is inelastic in the long-run, these future shocks raise future rents, and prices rise today to reflect this fact. This price change occurs even if supply is perfectly elastic today, because residents anticipate the near future in which supply will not be able to adjust as easily.

This supply condition—elastic short-run supply, inelastic long-run supply—occurs in cities at an intermediate level of development. Figure 2.5(a) demonstrates the possible combinations of short-run and long-run supply elasticities in a city. We plot the pass-through $1/(\epsilon^S + \epsilon^D)$; a higher pass-through corresponds to a lower elasticity. Lightly developed cities have highly elastic short-run and long-run supply, and heavily developed

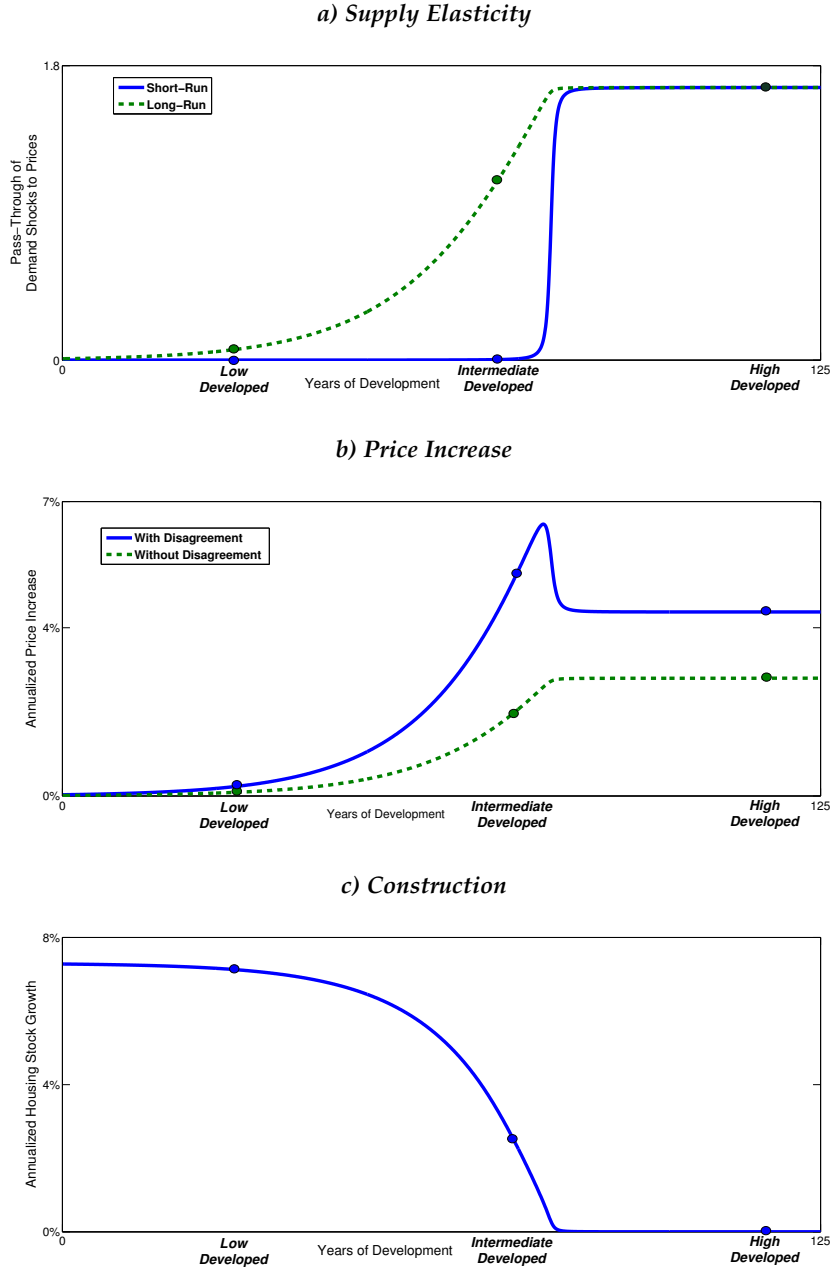


Figure 2.5: Model Simulations For Different Cities

Notes: The parameters we use are $\tilde{\mu} = 1$, $\bar{\mu} = 1$, $x = 0.06$, $g = 0.013$, $\epsilon^D = 1$, $\beta = 0.93^6$, and $\epsilon^I = 1$. We hold the amount of space S fixed and vary the initial trend demand given by $\log \bar{N}_0/g$. **(a)** Short-run pass-through is $1/(\epsilon_0^S + \epsilon^D)$; long-run pass-through is $1/(\tilde{\epsilon}^S + \epsilon^D)$. We calculate the rent and housing stock at each level of development using (A1) in the Appendix, and then calculate the supply elasticities using (2.4). **(b)** Each curve reports the derivative in (2.12) times x , which we calculate using the elasticities shown in panel (a). The “without disagreement” counterfactual uses $\tilde{\mu} = \bar{\mu} = 0.2$ instead of $\tilde{\mu} = 1 > \bar{\mu} = 0.2$. **(c)** We plot the construction equation (A2) using the elasticities shown in panel (a), as well as rents at each stage of development from (A1) and prices at each development stage from $p_0 = \sum_{t=0}^{\infty} \beta^t r_t^h$, which we calculate at $x = 0$.

cities have inelastic short-run and long-run supply. In the intermediate case, current supply is elastic while long-run supply is inelastic, reflecting the near future of constrained supply.

As we discussed in the Introduction, this theoretical supply condition describes the elastic markets that experienced large house price booms between 2000 and 2006. These cities found themselves in a state of arrested development as the boom began in 2000. Although ample land existed for current construction, long-run barriers constrain their future growth.

The second puzzle (2.12) explains why the price booms in these elastic cities were as large as those happening in inelastic cities at the same time. Disagreement amplifies the house price boom the most in exactly these nearly developed elastic cities. The amplification effect of disagreement equals the extent to which optimists bias the price increase given in (2.12). When owner-occupancy frictions are present ($\chi = 1$), the difference between the price boom under disagreement and under the counterfactual in which all residents hold the average belief $\bar{\mu}$ is

$$\frac{\epsilon_0^S}{\epsilon_0^S + \epsilon^D} \frac{\tilde{\mu} - \bar{\mu}}{\tilde{\epsilon}^S + \epsilon^D}.$$

This amplification is largest in nearly developed elastic cities, where ϵ_0^S is large and $\tilde{\epsilon}^S$ is small. Because this amplification increases in ϵ_0^S and decreases in $\tilde{\epsilon}^S$, nearly developed elastic cities provide the *ideal* condition for disagreement to amplify a house price boom. Implication 8, which we prove in the Appendix, states this result formally.

Implication 8 *Disagreement amplifies house price booms most in cities at an intermediate level of development, as long as owner-occupancy frictions are large enough. Define Δ to be the difference between the price boom given in (2.12) and the counterfactual in which all residents hold the average belief $\bar{\mu}$. Then there exists $\chi^* < 1$ such that for $\chi^* \leq \chi \leq 1$, Δ is strictly largest at an intermediate level of initial development $\bar{N}_0^* < \infty$.*

Figure 2.5(b) plots the house price boom given by (2.12) across different levels of city development, for both the case of disagreement and the case in which all residents hold the average belief. The amplification effect of disagreement is the difference between the two curves. Optimistic speculators amplify the price boom the most in the intermediate city.

Highly elastic short-run supply facilitates speculation in land markets, biasing prices towards their optimistic belief. This bias significantly increases house prices because housing supply is constrained in the near future. The optimism bias is smaller in the highly developed city. As a result, price increases in intermediate cities are as large as the price boom in the highly developed areas.

In fact, the price boom in some intermediate cities can *exceed* that in the highly developed cities. The parameters we use in Figure 2.5(b) generate an example of this phenomenon. This surprising result reverses the conclusion of standard models of housing cycles, in which the most constrained areas always experience the largest price increases. This reversal occurs as long as owner-occupancy frictions are high and the extent of disagreement is sufficiently large:

Implication 9 *If disagreement and owner-occupancy frictions are large enough, then the largest house price boom occurs in a city at an intermediate level of development. There exists $\chi^* < 1$ and $\delta > 0$ such that for $\chi^* \leq \chi \leq 1$ and $\tilde{\mu} - \bar{\mu} \geq \delta$, the price boom $d \log p_0^h / dx$ is strictly largest at an intermediate level of development $\bar{N}_0^* < \infty$.*

Our model has succeeded in explaining the large house price booms in the elastic cities without arguing that these cities experienced abnormally large housing demand shocks. These markets experienced some of the largest house price booms in the country because of their supply conditions, not in spite of them.

The final puzzle explained by (2.12) is why large house price booms occurred in some elastic cities but not in others. Elastic cities are those for which $\epsilon_0^S \approx \infty$. As shown in Figure 2.5(a), these cities differ in their long-run supply elasticity $\tilde{\epsilon}^S$. When $\tilde{\epsilon}^S = \infty$, the house price boom $d \log p_0^h / dx = 0$. Prices remain flat because construction can freely respond to demand shocks now and for the foreseeable future. House prices increase in elastic cities only when the development constraint will make construction difficult in the near future.

The elastic American cities which experienced stable house prices between 2000 and 2006 possess characteristics that lead long-run supply to be elastic. These cities, located in Texas and other central American areas, are characterized by flat geography, a lack of future

regulation, and homogeneous sprawl (Glaeser and Kahn, 2004; Glaeser and Kohlhase, 2004; Burchfield et al., 2006; Glaeser, Gyourko and Saiz, 2008; Saiz, 2010). These conditions allow the cities to expand indefinitely, leading S to be infinite or very high. Unlimited land leads the elasticity of supply to remain infinite forever, according to (2.4).

The level of house prices before the shock identifies the difference between the elastic cities that can expand indefinitely and elastic cities that face constraints in the near future. House prices increase with development. Therefore, the elastic cities nearing their development constraints should have higher house prices before the shock than the other elastic cities. The following implication summarizes these results.

Implication 10 *Consider two cities that experience the same demand shock and in which current housing supply is perfectly elastic ($\epsilon_0^S = \infty$). House prices rise more in the city in which the long-run supply elasticity $\tilde{\epsilon}^S$ is lower. Before the shock occurs, a greater share of the land in this city is already developed, and the level of house prices is higher.*

In practice, calculating a metro-area house price level is difficult because characteristics such as construction costs vary widely within and across metro areas, although valiant attempts have been made (Glaeser and Gyourko, 2005; Davis and Heathcote, 2007; Nichols, Oliner and Mulhall, 2013). With the appropriate data, we would be able to distinguish the low-developed from the medium-developed cities.

We have used (2.12) to explain the large house price increases in certain elastic housing markets in the U.S. between 2000 and 2006. An additional salient feature of these booms is that they coincided with rapid construction. As we document in Section 2.2, these cities experienced some of the most intense permitting activity in the nation during this period. Our model captures this phenomenon. Figure 2.5(c) plots the construction response to the shock in different cities. In cities where current housing supply is elastic, new construction accommodates the shock. The elastic cities include both the lightly developed and intermediate developed areas.

2.6 Variation in House Price Booms Within Cities

The model also makes predictions on the variation in house price increases within a given city. Optimistic speculators hold rental housing, just as they hold land. Prices appear more optimistic, and hence house price booms are larger, in market segments where a greater share of housing is rented.

This result emerges from (2.12). Recall that χ is the share of the housing stock that is owner-occupied rather than rented when $x = 0$. It is a sufficient statistic for the distribution F_a of owner-occupancy utility. When χ is larger, the price increase $d \log p_0^h / dx$ is smaller:

$$\frac{\partial}{\partial \chi} \frac{d \log p_0^h}{dx} = -\frac{\epsilon^D}{\epsilon_0^S + \epsilon^D} \frac{\tilde{\mu} - \bar{\mu}}{\tilde{\epsilon}^S + \epsilon^D} < 0.$$

This derivative is negative because the optimistic belief $\tilde{\mu}$ exceeds the average belief $\bar{\mu}$.

A city's housing market consists of a number of market segments, which are subsets of the housing market that attract distinct populations of residents. Because they attract distinct populations, we can analyze them using (2.12), which was formulated at the city-level. All else equal, housing submarkets in which χ is higher experience smaller house price booms:

Implication 11 *Suppose market segments within a city differ only in χ , the relative share of renters versus owner-occupants they attract: the shock x and the short-run and long-run supply elasticities ϵ_0^S and $\tilde{\epsilon}^S$ are constant within a city. Then house price booms are smaller in market segments where χ is larger.*

2.6.1 Location

We first consider variation in χ across neighborhoods. Neighborhoods provide an example of market segments because they differ in the amenities they offer. For instance, some areas offer proximity to restaurants and nightlife; others are characterized by access to good public schools. These amenities appeal differentially to different populations of residents. Variation in amenities hence leads χ to vary across space. Neighborhoods whose amenities appeal relatively more to owner-occupants (high a residents) than to renters (low a residents)

are characterized by a higher value of χ .

Consistent with Implication 11, house prices increased more between 2000 and 2006 in neighborhoods where χ was higher in 2000. We obtain ZIP-level data on χ from the U.S. Census, which reports the share of occupied housing that is owner-occupied, as opposed to rented, in each ZIP code in 2000. The fraction χ varies considerably within cities. Its national mean is 0.71 and standard deviation is 0.17, while the R^2 of regressing χ on city fixed-effects is only 0.12. We calculate the real increase in house prices from 2000 to 2006 using Zillow.com's ZIP-level house price indices. We regress this price increase on χ and city fixed-effects, and find a negative and highly significant coefficient of -0.10 (0.026), where the standard error is clustered at the city level.

However, this negative relationship between χ and price increases may not be causal. Housing demand shocks in this boom were larger in neighborhoods with a lower value of χ . The housing boom resulted from an expansion of credit to low-income households (Mian and Sufi, 2009; Landvoigt, Piazzesi and Schneider, 2013), and ZIP-level income strongly covaries with χ .¹⁰

The appeal of χ is that it predicts price increases in any housing boom in which there is disagreement about future fundamentals. In general, χ predicts price increases because it is negatively correlated with speculation, not because it is correlated with demand shocks. Empirical work can test Implication 11 by examining housing booms in which the shocks are independent from χ .

2.6.2 Structure Type

The second approach to measuring χ is to exploit variation across different types of housing structures. According to the U.S. Census, 87% of occupied detached single-family houses in 2000 were owner-occupied rather than rented. In contrast, only 14% of occupied multifamily housing was owner-occupied. According to Implication 11, the enormous difference in χ

¹⁰The IRS reports the median adjusted gross income at the ZIP level. We take out city-level means, and the resulting correlation with χ is 0.40.

between these two types of housing causes a larger price boom in multifamily housing, all else equal.

This result squares with accounts of heightened investment activity in multifamily housing during the boom.¹¹ For instance, a consortium of investors—including the Church of England and California’s pension fund CalPERS—purchased Stuyvesant Town & Peter Cooper Village, Manhattan’s largest apartment complex, for a record price of \$5.4 billion in 2006. Their investment went into foreclosure in 2010 as the price of this complex sharply fell (Segel et al., 2011). Multifamily housing attracts speculators because it is easier to rent out than single-family housing. Optimistic speculators bid up multifamily house prices and cause large price booms in this submarket during periods of uncertainty.

2.7 Conclusion

In this paper, we argue that speculation explains an important part of housing cycles. Speculation amplifies house price booms by biasing prices toward optimistic valuations. We document the central importance of land price increases for explaining the U.S. house price boom between 2000 and 2006. These land price increases resulted from speculation directly in the land market. Consistent with this theory, homebuilders significantly increased their land investments during the boom and then suffered large capital losses during the bust. Many investors disagreed with this optimistic behavior and short-sold homebuilder equity as the homebuilders were purchasing land.

Our emphasis on speculation allows us to explain aspects of the boom that are at odds with existing theories of house prices. Many of the largest price increases occurred in cities that were able to build new houses quickly. This fact poses a problem for theories that stress inelastic housing supply as the source of house price booms. But it sits well with our theory, which instead emphasizes speculation. Undeveloped land facilitates speculation due to rental frictions in the housing market. In our model, large price booms occur in elastic cities

¹¹Bayer, Geissler and Roberts (2013) develop a method to identify speculators in the data. A relevant extension of their work would be to look at the types of housing speculators invest in.

facing a development barrier in the near future—cities in arrested development.

Our approach also makes some new predictions. Price booms are larger in submarkets within a city where a greater share of housing is rented. Although we presented some evidence for this prediction, further empirical work is needed to test it more carefully.

In all, we have presented a different but complementary story of the sources of housing cycles than the literature has offered. Our theory explains several puzzles and suggests new directions for empirical research.

Chapter 3

Regulators vs. Zombies: Loss Overhang and Lending in a Long Slump

3.1 Introduction

Why are economies slow to recover from recessions that coincide with financial crises? Reinhart and Rogoff (2009) document the recurrence of this pattern in macroeconomic data going back to the thirteenth century. Its latest return followed the U.S. financial crisis of 2008, in what Hall (2011) refers to as the Great Slump. The fact has inspired much debate among researchers and policymakers.

This paper presents micro-level evidence of one possible mechanism responsible for the persistence of long slumps. The banking sector suffers from *loss overhang*, in which the need to write down losses in the future against impaired loans on the balance sheet today reduces the value of making new, positive NPV loans. The standard debt overhang mechanism of Myers (1977) is part of this story, since a suffering bank with unmarked losses would be severely undercapitalized were those losses recognized. But an additional constraint comes from bank incentives to deter regulatory action. When a healthy loan matures and pays

off, a bank can improve its risk capital ratio by investing the proceeds in less risky assets, such as deposits in other banks. This insight helps clarify the ambiguous predictions of the debt overhang model about a bank's incentive to take risks when near insolvency. I refer to banks afflicted with loss overhang as zombie banks, or *zombies*.

I use a sample of several hundred FDIC-induced resolutions during the Great Slump to show how loss overhang impairs and distorts lending prior to a zombie's failure. The FDIC and its companion regulators have the legal right to take banks into receivership in order to protect the claims of depositors. Remaining creditors and shareholders take losses while the FDIC arranges the assumption of insured deposits and the sale of underlying assets. In the process, what remains of the bank's balance sheet is refreshed, with future losses being separated from rolled-over deposit funding.

Prior to failure, zombies operate with a significant portion of their loan book marked well above market value. Subsequent to resolution, the FDIC sells these loans in the secondary market at a discount between 40 and 70 percent of book value. This loss overhang restricts lending in relatively healthy loan categories and leads zombies to hoard safe assets. Thus, loss overhang may also offer a micro-level theory of excess savings in the liquidity trap that is directly related to ongoing problems in the banking sector.

Because I cannot observe the zombies in isolation after resolution, my main empirical strategy involves generating synthetic banks by merging the records for zombie and eventual acquirer banks and assessing the treatment effect of resolution on lending activity in these synthetic banks. The FDIC is most likely to close banks that are in the worst shape, so post resolution growth is probably lower than if zombies were treated at random. This also implies that any selection bias should work against finding a resumption in lending post resolution.

I use zombie characteristics and losing bidder banks to address concerns about omitted acquirer traits driving my results. When the zombie bank is large relative to the synthetic bank, the reversal of pre-failure trends in lending is larger. Thus acquirer activity post resolution depends on the new funding from the acquired bank. A second approach uses

the post failure lending paths for losing bidder banks to generate counterfactual lending paths for the acquiring bank. I combine this counterfactual path with the true acquirer's post failure path to infer the counterfactual lending path for the zombie bank alone. Again I find that lending resumes for zombies following resolution.

As the crisis evolved, the FDIC may have run into resource constraints—in both liquidity and manpower—that limited its ability to deal with the growing number of zombies. The number of banks operating under regulatory notice soared relative to the number of banks being closed. As a consequence, a substantial number of banks were left in a stultifying zombie state for months and perhaps years. In the first quarter of 2010, one in eight commercial banks were classified as problem banks by the FDIC.

Not only does loss overhang restrict lending in these surviving zombies, healthy banks in the same area appear unable to pick up the slack. The result is slower growth. Most zombies are small to medium-sized banks suffering from losses in residential and commercial construction and land development loans. These losses appear to spill over into unrelated sectors, such as services, mining, and manufacturing. Employment in zombie counties—counties with a high fraction of local deposits in zombie banks—recovers more slowly in these sectors.

The label zombies carries with it the suggestion that the only appropriate policy for healing these banks is to kill them, the goal being to allow the zombies to shed their bad assets. In 2008, policymakers dealt with struggling banks mainly by recapitalizing them, rather than working to remove impaired loans from bank balance sheets. While both policies might have the same direct effect on bank capital ratios, the latter may still leave banks with an expected future capital loss. Thus, my evidence on loss overhang suggests a reason why a good bank/bad bank resolution may be preferred to direct equity injections, if speeding the resumption of lending is an objective.

Section 3.2 places this paper in the context of related work. Section 3.3 provides an overview of the pertinent U.S. banking regulations and the bank resolution process. Section 3.4 introduces the data. Section 3.5 introduces the synthetic bank series, analyzes these

series around the failure event, and documents substantial discounts on zombie loan pools. Section 3.6 explores the effect of FDIC-induced failure on lending relative to the universe of banks, to surviving zombies, and to counterfactual lending paths generated from losing bidders. Section 3.7 explores how the county-level incidence of deposits in zombie banks affects growth. Section 3.8 concludes.

3.2 Related Literature

This study has a conceptual analogue in Bernanke (1983), who argues that a credit crunch induced by bank failures severely exacerbated the Great Depression. However, data availability limits his ability to identify the mechanism through which bank balance sheets impede growth.

The motivating macroeconomic facts follow from the macro empirical literature on economic recoveries following financial crises, typified by Reinhart and Rogoff (2009). Using cross-country and mainly informal historical analyses, these papers emphasize the persistent weakness in asset and labor markets and dramatic increases in government indebtedness following financial crises. Empirical study stays at the level of slow-moving, macroeconomic aggregates. As a consequence the mechanism through which financial contraction frustrates recovery and the appropriate policy response remain unclear,

This paper makes a complementary argument to those in Peek and Rosengren (2000) and Ashcraft (2005), in which weakening or closing a healthy bank worsens local economic outcomes. Here, closing a *sick* bank *improves* local economic outcomes.

This paper also contributes to the literature on the bank lending channel spawned by Bernanke and Blinder (1992) and assessed empirically by Kashyap and Stein (2000). The theoretical foundation for these papers relies on contractions in bank balance sheets caused by monetary policy shocks. Unlike these papers, my focus is on balance sheets impaired by unrecorded or anticipated losses rather than policy-induced reductions in bank reserves. In a sense, I am applying the mechanism from Bernanke and Gertler (1989), which focuses on borrower balance sheets, to understand how lender balance sheets can influence the

economy through the bank lending channel.

My findings relate to those in Mian and Sufi (2009) and Mian and Sufi (2012), who show how household indebtedness led to reduced demand in some parts of the country during the Great Slump. By documenting an alternative source of reduced demand—constrained supply of new loans by local banks—I show how demand may have impeded growth in those regions that did not see dramatic expansions in subprime lending. Mian and Sufi (2012) focus on non-tradable goods, while I show how this force may have restricted employment growth in tradable goods industries.

Caballero, Hoshi and Kashyap (2008) show how struggling banks misdirected loans to otherwise insolvent, zombie firms and how this “evergreening” distorted competition in product markets. In the Great Slump, American bank regulators anticipated this kind of activity and used regulatory agreements and bank examinations to try and prevent it. For example, most agreements include a clause proscribing the extension of credit to borrowers not current on existing loans. In addition, while the Japanese government encouraged banks to keep lending through its crisis, the American government appears to have been more concerned with preserving financial stability.¹

3.3 Overview of U.S. Banking Regulation

The key institutional details relevant to this study concern how regulators monitor and control struggling banks and how the FDIC administers resolution of failed banks.

3.3.1 Prompt Corrective Action

Standards for adequate bank capitalization were legislated in the Federal Deposit Insurance Act (FDIA). When a bank is not adequately capitalized, the FDIA requires regulators to take “prompt corrective action” to address the situation. The act established five capital tiers: “well capitalized,” “adequately capitalized,” “undercapitalized,” “significantly undercapitalized,”

¹See Hoshi and Kashyap (2010) for an elaboration of this difference.

and “critically undercapitalized.” Banks are classified based on standards for the total capital ratio, the Tier 1 capital ratio, and the leverage ratio. For example, a “well capitalized” bank has a total risk-based capital ratio of at least 10 percent, a Tier 1 ratio of at least 6 percent and a leverage ratio of at least 5 percent. A “critically undercapitalized” bank has tangible equity equal to or less than 2 percent of average quarterly tangible assets.

Undercapitalized institutions are subject to growth limitations and required to submit a capital restoration plan. Additional requirements for “significantly” and more severely undercapitalized banks include requirements to issue additional stock, to reduce total assets, and to stop taking deposits from other banks. When a bank becomes “critically undercapitalized,” the FDIA requires a receiver to be appointed within 90 days, unless another documented path to recovery has been identified. Section 38 of the FDIA authorizes the FDIC to delay mandatory resolution for an additional 180 days to prevent risk of loss to the deposit insurance fund.

3.3.2 FDIC Resolutions

The FDIC² has two roles in every bank failure: to protect the depositors in the amount of their insured deposits and to administer the receivership for the failed institution following resolution. The chartering authority for a bank typically closes it when it has become insolvent, critically undercapitalized, or unable to meet requests for deposit withdrawals.

There are three basic resolution methods: purchase and assumption transactions, deposit payoffs, and open bank assistance transactions.³ Activities begin with the receipt of a Failing Bank Letter, after which the FDIC sends in a team of specialists to assemble an information package (including a statistical sampling-based asset valuation review). The FDIC determines all possible resolution methods and moves to market the failing bank by

²This section draws on the FDIC resolutions handbook, available at <http://www.fdic.gov/bank/historical/reshandbook/index.html>. It describes each FDIC strategy for dealing with a failed bank and outlines receivership operating protocols. The handbook was originally written to collect and preserve the wisdom accumulated following the savings and loan crisis, so it is possible that some of these details are stale.

³Not used since 1992.

inviting bidders and providing them with the information package.

Bids are submitted as (1) a premium for the franchise value of the failed bank's deposits and (2) a bid for all or part of the institution's assets. Since 1991, the FDIC has been required to choose the resolution alternative (including bid) that is least costly to the deposit insurance fund. The FDIC transfers cash to an acquiring institution in an amount equal to liabilities assumed less assets purchased less franchise premium. The entire resolution process takes 90 to 100 days.

Purchase and assumption (P&A) transactions are the most common resolution method. Within this category, there are variations as to how assets are transferred. These variations include loan purchase, put option, asset pool, whole bank, loss sharing, and bridge bank P&As. In a loss sharing transaction, the FDIC absorbs typically 80 percent of credit losses, usually on business and commercial real estate loans.

In basic P&As, assets passing to acquirers are generally limited to cash and cash equivalents. These were more common prior to the codification of bidder due diligence practices in the early nineties. In loan purchase P&As, the winning bidder assumes a small portion of the portfolio, usually only the installment loans, more generally between 10 and 25 percent of assets. A modified P&A is a loan purchase P&A plus some of the mortgage loan portfolio, more generally between 25 and 50 percent of failed bank assets. Optional asset pool P&As allow the FDIC to market homogeneous loan pools to other acquirers. Whole bank P&As transfer all assets, but at a substantial discount. Loss sharing P&As transfer assets, but not at a premium. Instead, cash moves to the acquirer as charges are taken and from the acquirer as recoveries are realized. During the S&L period in the early nineties, as the crisis became more acute, the FDIC moved towards resolutions that passed more of the assets to the acquirer.

3.4 Data

I assemble data from several sources, briefly described here. Bank balance sheet data for the years 2001 through 2011 come from the FFIEC quarterly Consolidated Reports of

Condition and Income (Forms 031/041), informally referred to as the Call Reports. Data on the deposits and geographical locations of bank branches come from the FDIC's annual Summary of Deposits. For measuring local employment, payrolls, and establishment counts at the county level, I use the Census Bureau's quarterly Census of Employment and Wages.

I draw a variety of items from the FDIC's web site. These include the identity and dates of bank failures, the names of acquiring banks, the names of losing bidder banks, estimated losses from resolution, and transaction details for sales of loan pools out of receivership. In addition, I draw details on the Deposit Insurance Fund from the Chief Financial Officer's quarterly reports to the FDIC's board.

There is no official public record of problem, or zombie, banks. The FDIC's Quarterly Banking Profile only reports aggregate statistics on banks with CAMELS ratings of three, four, or five. However, CalculatedRisk, an economics web log, assembles an unofficial record of all banks operating under Prompt Corrective Actions and Cease and Desist orders. Regulators are required to publish these agreements. I use the unofficial list to identify zombie banks beginning in 2007 and running through 2011. The unofficial list matches the official numbers closely.

3.5 Bank Failures in the Great Slump

Figure 3.1 documents the sample of bank deletions studied in this paper. I focus on FDIC-assisted mergers because those are the bank failures for which we can observe post-failure lending activity. Assisted mergers account for the majority of failures during this period. The crisis episode begins in 2008 with 44 failures that year, half of which were assisted mergers. In 2009, 133 banks failed with 114 of those being assisted mergers. In 2010, there were 130 assisted mergers accounting for all of the assisted failures. During this time, a number of unassisted mergers occurred as well, though at a slower rate than during the expansion. In 2008, there were 260 unassisted mergers. In 2009 and 2010, there were 158 and 183 unassisted mergers, respectively.

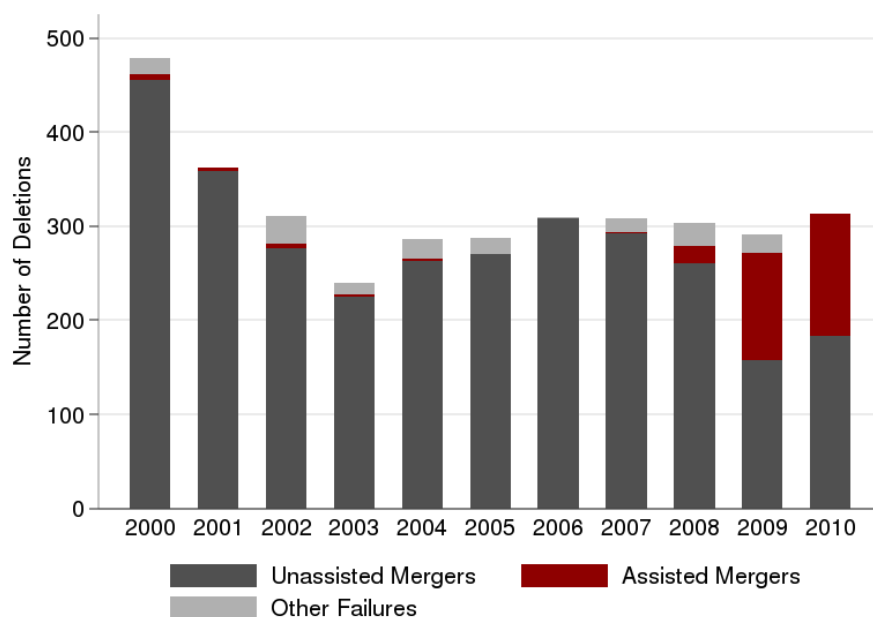


Figure 3.1: *Bank Deletions since 2000*

Notes: Data come from the FDIC CB tables.

For each zombie,⁴ I generate synthetic bank time series by combining the quarterly Call Reports for the zombie prior to closure with the Call Reports for the acquiring bank. I then match these synthetic series to the Call Reports for the acquiring bank after the closing date. The result is a record of assets, liabilities, loans, and losses in a hypothetical bank that can be treated with FDIC resolution and compared to the population using a standard differences-in-differences (DD) approach.

Figure 3.2 displays these synthetic time series along with the pre closure, zombie time series, shown in event time around the failure date. For visual convenience, I normalize each series to equal one six quarters prior to failure. On average, 20 percent of synthetic bank loans pass from the zombie's balance sheet to the FDIC or to remarked loss-share assets.⁵

⁴I use the term *zombie* to refer to any bank operating with potential losses large enough to induce insolvency. Here, these are banks deemed by regulators to be too weak to recover without an FDIC-assisted merger.

⁵Loss-share assets are recorded in a separate field in the Call Reports. Because the FDIC bears 80 percent of the risk of these assets, the acquiring institution can carry them at a maximum of 20 percent of prior book value.

Loans staying with the receivership are concentrated in the commercial and industrial (C&I), construction, and commercial real estate (CRE) categories, indicated by the sharp drops in Figure 3.2. This motivates my analysis of lending activity in these categories following an FDIC-assisted resolution.

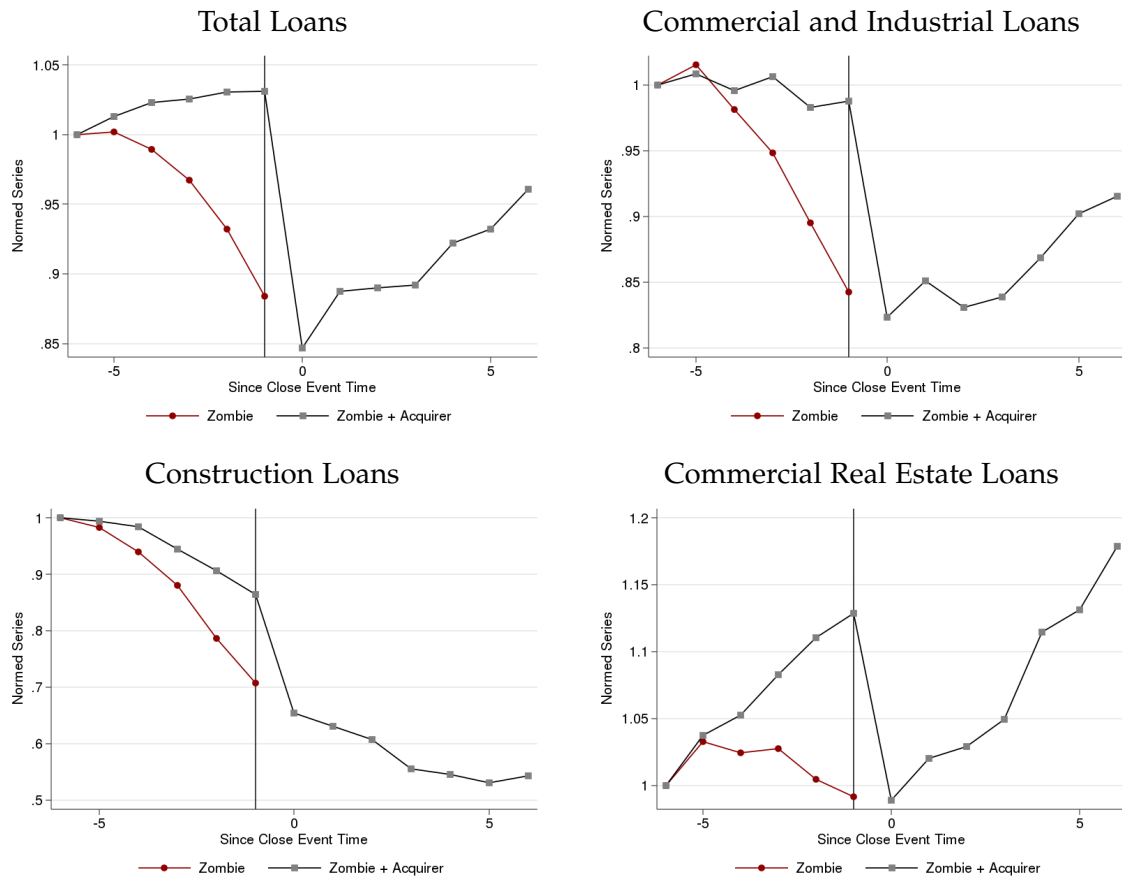


Figure 3.2: *Loan Paths for Zombie and Synthetic Banks*

Notes: The x-axis tracks event time by quarter around the date a bank is absorbed into an acquiring bank, with time 0 being the first call report date after the event. Variables are normalized by their value 6 quarters prior to the event date.

Total lending in the average synthetic bank grows much more slowly in the pre period and actually contracts in the case of C&I loans. Sharp contractions in these categories for the zombie banks drive this pattern. The paths for construction and CRE loans are similar in the pre and post periods.

Figure 3.3 provides evidence that loans passing from zombies into receivership are

substantially impaired. The FDIC web site offers details about loan pools which it has sold out of receivership. Using this data, I compute the typical discount to book value applied to these pools upon sale. Non-performing loans take a 60 percent discount on average, while even performing loans appear to take a 30 percent discount. Lost relationship capital appears insufficient to explain these discounts, since non-performing loans presumably carry less valuable relationship capital when on the originator's balance sheet. Further anecdotal evidence that zombies carry loans at generous book values comes directly from the FDIC's resolution handbook, which states in passing that "asset values are generally overstated in a failing bank or thrift."

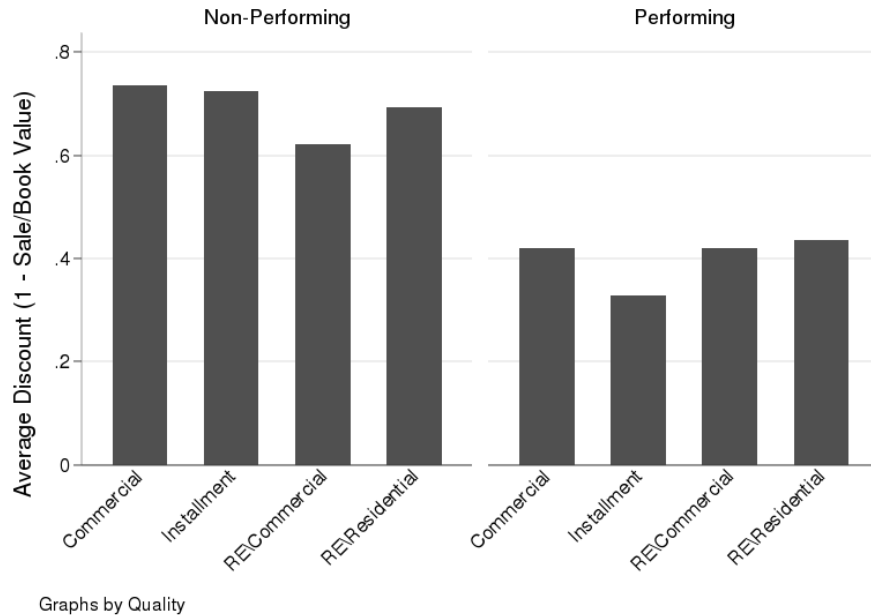


Figure 3.3: *Loan Sale Discounts by Loan Category*

Notes: Data from the FDIC on loan pool sales since 2005. The discount is computed relative to book value reported by the FDIC.

More direct evidence of bad loans passing out of the banking system comes from the synthetic time series. Figure 3.4 plots the seriously delinquent and non-accrual loan shares across loan categories around the failure date. As a fraction of bank assets, residential and commercial construction loans are the most impaired with more than three percent of the

synthetic bank's loan book in this category pre failure. This share falls by 50 percent in the quarter immediately following resolution. The CRE and C&I non-performing loan shares also fall after resolution.

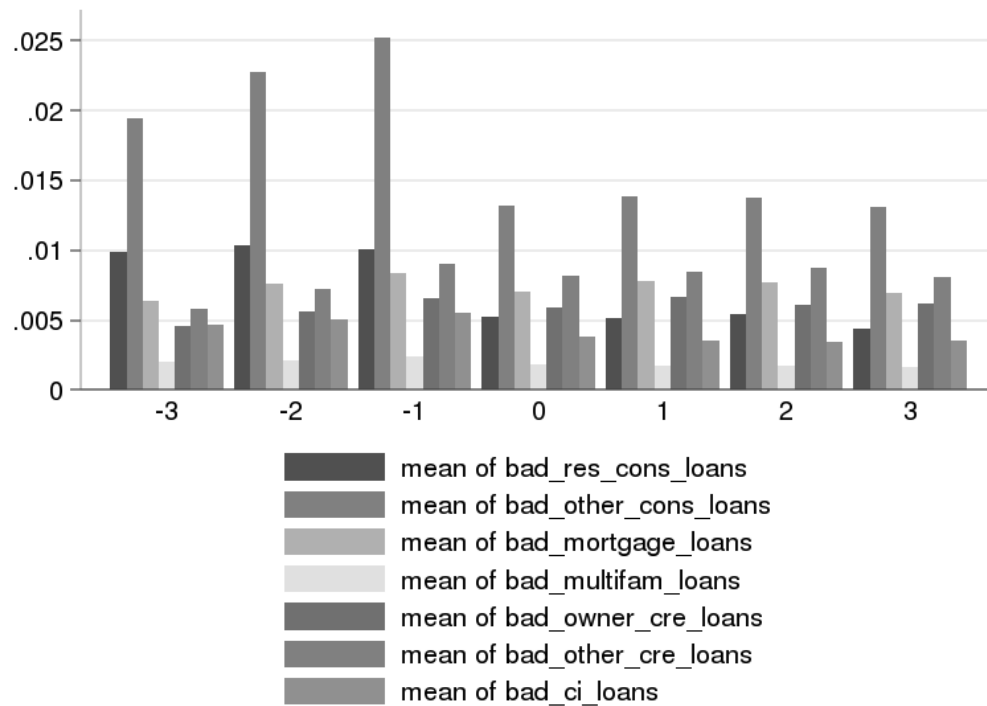


Figure 3.4: Problem Loan Shares for Synthetic Banks

Notes: The x-axis tracks event time by quarter around the date a bank is absorbed into an acquiring bank, with time 0 being the first call report date after the event. Shares are computed by adding the non-accrual and 90+ days delinquent bins and dividing by total loans.

Figure 3.5 tracks bank capital, bank deposit holdings in other banks, charge-offs, and the loan loss provision balance for the average zombie and synthetic banks. The average bank is nearly insolvent at the time of failure, reporting an equity ratio of two percent in its final Call Report. Leading to failure, zombie holdings of deposits in other depository institutions triple. Taken with the evidence on lending in Figure 3.2, zombies appear not to be engaged in risk-shifting behavior but rather to be stockpiling safe assets in an attempt to deter regulatory action. Of course, this pattern may be a mechanical consequence of

increased loan loss provisioning. Were this the case, we would expect the same pattern to appear for the acquiring bank, as loan loss provisions do trend up following the acquisition. However, acquiring bank holdings of other bank deposits only increase slightly post merger.

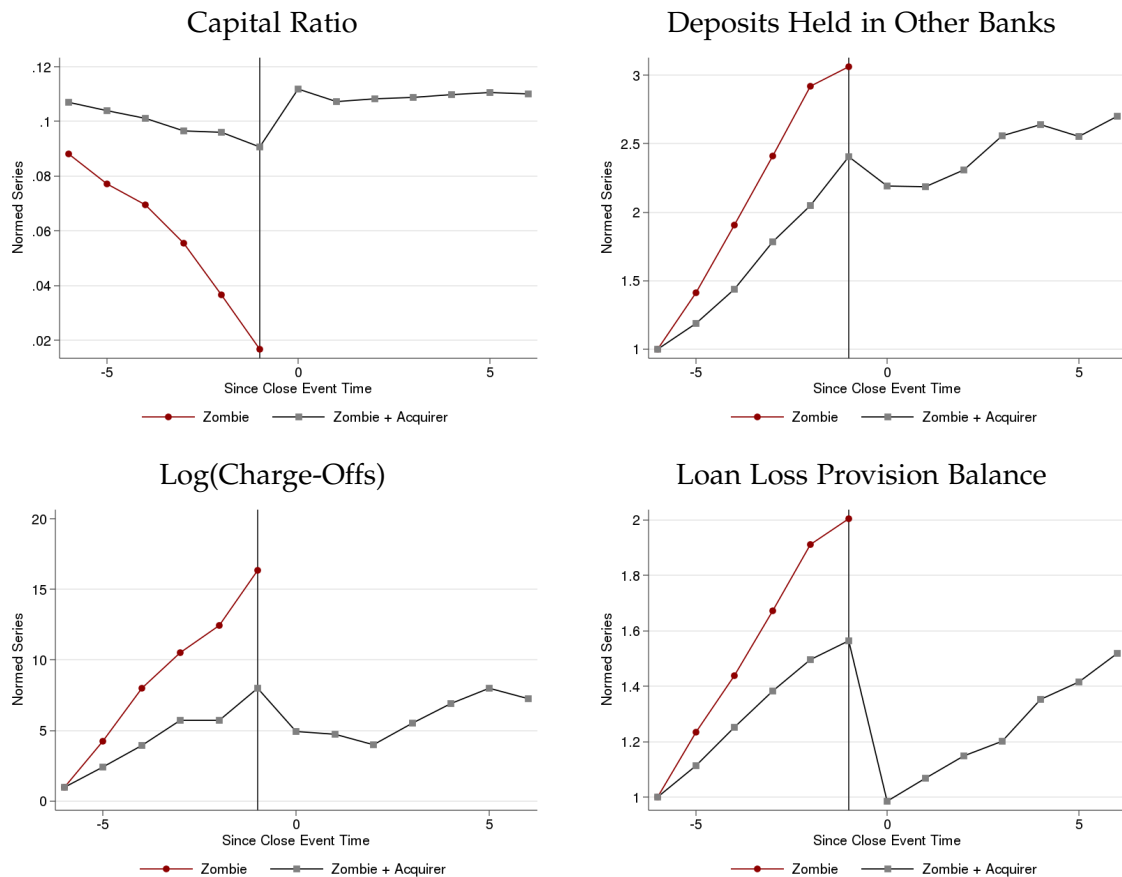


Figure 3.5: Capital and Loss Paths for Zombie and Synthetic Banks

Notes: The x-axis tracks event time by quarter around the date a bank is absorbed into an acquiring bank, with time 0 being the first call report date after the event. Variables are normalized by their value 6 quarters prior to the event date.

Table 3.1 compares the three groups of banks in this study: synthetic banks, zombie banks, and the remaining commercial banks. I mark a bank as a zombie if it is operating under a PCA or Cease and Desist agreement with its regulator. My sample of synthetic banks includes 252 failed bank/acquirer pairs. In Table 3.1, we can see that zombie banks are less capitalized, have assets allocated more toward loans, and their loans are more concentrated in construction and CRE. These patterns come through in the synthetic column,

though the median zombie is one fifth the size of its synthetic counterpart.

3.6 Resurrected Lending following Resolution

The ideal experiment for testing the loss overhang channel would randomly assign a group of zombies to have their bad loans swapped for cash at face value. Comparing the lending paths and asset allocations pre and post treatment to a group of control zombies would allow me to measure the causal effect of this swap and to infer how losses restrict and distort the financing of new projects. I map this ideal to the data by studying FDIC-assisted mergers and the paths of synthetic zombie-acquirer pairs around the merger event.

FDIC-assisted mergers allow zombies to shed much of their current and future non-performing loan portfolios by absorbing these assets into an FDIC-operated receivership. Thus, they provide the closest empirical analog to a hypothetical bad loan swap. The 266 assisted mergers that occurred between 2007 and 2010 provide the largest sample of such events in the U.S. since the savings and loan crisis.

These resolutions are not a perfect test because I can only observe the zombie's activity prior to resolution. As a result, I focus on the synthetic balance sheets created by combining the zombies with their eventual acquirers. The design is a generalized difference-in-differences (DD) approach, in which FDIC resolution is the treatment variable. The universe of healthy banks and the set of surviving zombies act as alternative control groups.

3.6.1 Results from the Basic Regression

Tables 3.2 and 3.3 present panel regressions of quarterly growth rates on a treatment indicator for FDIC-assisted merger and a series of controls, including the mean bank size, year-by-quarter fixed effects, and state fixed effects. I exclude the quarter immediately following resolution in order to avoid counting mechanical changes occurring during the merger as real activity. Because the left hand variable is a first difference, I do not include bank fixed effects, under the assumption that the error term in first differences is independent of time-invariant factors. Bertrand, Duflo and Mullainathan (2004) warn that serial correlation

Table 3.1: Bank Statistics during the Great Slump**(a) First Quarter 2008**

Bank Type	Synthetic	Zombie	Universe
Assets (Median, 000s)	1,043,870	190,434	128,816
Assets (000s)	7,421,626	441,919	1,513,299
Total Loans/Assets	0.724	0.737	0.652
Deposits in Other Banks/Assets	0.030	0.031	0.048
Book Equity/Assets	0.111	0.098	0.132
Construction Loans/Total Loans	0.219	0.304	0.122
CRE Loans/Total Loans	0.288	0.255	0.237
Count	252	165	7664

(b) First Quarter 2009

Bank Type	Synthetic	Zombie	Universe
Assets (Median, 000s)	1,097,194	165,751	139,036
Assets (000s)	8,311,734	427,474	1,669,090
Total Loans/Assets	0.702	0.711	0.658
Deposits in Other Banks/Assets	0.051	0.068	0.067
Book Equity/Assets	0.099	0.076	0.115
Construction Loans/Total Loans	0.182	0.196	0.102
CRE Loans/Total Loans	0.306	0.276	0.247
Count	252	299	7448

(c) First Quarter 2010

Bank Type	Synthetic	Zombie	Universe
Assets (Median, 000s)	1,091,840	193,036	148,149
Assets (000s)	8,865,004	451,767	1,738,028
Total Loans/Assets	0.662	0.691	0.637
Deposits in Other Banks/Assets	0.086	0.089	0.085
Book Equity/Assets	0.100	0.075	0.110
Construction Loans/Total Loans	0.127	0.139	0.079
CRE Loans/Total Loans	0.334	0.332	0.256
Count	246	674	7178

Notes: Data come from quarterly Consolidated Reports of Condition and Income (Call Reports). Synthetic banks are the consolidated Call Reports for failed banks and their prospective acquirers. Zombie banks are banks operating under an active Prompt Corrective Action or Cease & Desist notice. Construction loans combine residential and non-residential construction and land development loans. CRE loans combine owner-occupied and non-owner-occupied commercial real estate loans.

in DD designs can severely bias inference toward finding an effect. As a result, I cluster standard errors at the bank level.

The average zombie is 20 percent the size of the combined entity, so the actual treatment effect will be significantly diluted by being pooled with a healthy bank's outcome path. Despite this bias against my experimental design, Tables 3.2 and 3.3 show that resolution leads to a substantial recovery in lending in the synthetic bank. When compared to the full universe of healthy banks in Table 3.2, the post period shows synthetic lending expanding in the post period by between 1.4 and 3.9 log points on a quarterly basis. This is a substantial recovery, considering the pre period growth rate for synthetics is 0.4 log points, while the system-wide lending growth rate is 2.8 log points over this period. This recovery is most evident and robust for C&I loans, where lending growth rates increase between 1.4 and 3.4 log points for synthetics. Construction and CRE loans show a lending recovery relative to the universe only in some specifications.

When compared to the sample of surviving zombies in Table 3.3, the change in lending growth is even more dramatic. The post period increase in total lending is between 3.9 and 7.0 log points and the increase in C&I is between 5.0 and 7.4 log points. This is equivalent to a relative annual increase in total lending growth of between 15.6 and 28.0 log points. CRE lending increases substantially, between 2.0 and 3.5 log points per quarter, and construction lending contracts significantly less quickly with a relative increase of 4.9 to 5.8 log points.

Table 3.2: Lending and Asset Allocation Post Resolution (Synthetic Banks vs. Full Bank Universe)

	Total Loans			C&I Loans			Construction Loans		
Post Close	0.014** [3.44]	0.025** [6.10]	0.039** [3.15]	0.014** [2.83]	0.023** [4.37]	0.034* [2.23]	-0.0052 [-0.78]	0.01 [1.54]	0.042** [3.31]
Bank Size		-0.0053** [-9.38]	-0.0074** [-11.6]		-0.0039** [-5.35]	-0.0064** [-8.25]		-0.0074** [-7.91]	-0.0091** [-9.28]
Constant	0.027** [35.4]	0.089** [12.8]	0.12** [12.8]	0.020** [25.2]	0.066** [7.52]	0.094** [8.95]	-0.015** [-14.3]	0.073** [6.28]	0.12** [8.51]
Year-Qtr Effects State Effects	X	X	X	X	X	X	X	X	X
N	151,010	151,010	149,238	147,782	147,782	146,096	93,199	93,199	91,966
R ²	0.011	0.013	0.026	0.008	0.008	0.012	0.008	0.009	0.011

	CRE Loans			Deposits Held in Other Banks			Loan Loss Provision Balance		
Post Close	0.014* [2.32]	0.018** [3.00]	0.0066 [0.66]	-0.035** [-3.43]	-0.034** [-3.28]	-0.065 [-1.72]	0.044** [3.02]	0.038* [2.53]	0.086 [1.61]
Bank Size		-0.0020** [-3.46]	-0.0045** [-7.14]		-0.00047 [-0.74]	-0.0019** [-2.76]		0.0030** [5.27]	0.00013 [0.22]
Constant	0.024** [39.5]	0.048** [6.79]	0.079** [8.93]	0.053** [77.9]	0.059** [7.87]	0.074** [8.00]	0.039** [59.3]	0.0039 [0.57]	0.038** [4.63]
Year-Qtr Effects State Effects	X	X	X	X	X	X	X	X	X
N	100,401	100,401	99,036	144,340	144,340	142,584	143,228	143,228	141,583
R ²	0.007	0.007	0.012	0.021	0.021	0.022	0.008	0.008	0.019

Notes: These regressions track changes in key bank balance sheet variables around FDIC-assisted merger for the synthetic bank, comprising the problem bank and the acquiring bank, and the universe of healthy banks. The left hand side variable in each regression is the log change of the variable at a quarterly frequency. Bank size is the mean of log assets for each bank. For synthetic banks, I include observations for the two years before and after the merger, excluding the period immediately after the merger. Standard errors are clustered at the bank level. T-statistics are displayed in the brackets. The significance levels are: ** p<0.01, * p<0.05.

Table 3.3: Lending and Asset Allocation Post Resolution (Synthetic Banks vs. Surviving Zombies)

	Total Loans			C&I Loans			Construction Loans		
Post Close	0.039** [9.45]	0.059** [11.7]	0.070** [5.24]	0.050** [8.76]	0.064** [10.3]	0.074** [4.55]	0.050** [6.56]	0.049** [6.29]	0.058** [4.32]
Bank Size		-0.010** [-6.93]	-0.011** [-7.47]		-0.0077** [-5.03]	-0.0086** [-5.17]		0.00034 [0.21]	-0.00033 [-0.19]
Constant	0.023** [11.7]	0.15** [7.63]	0.15** [7.96]	0.011** [5.05]	0.10** [5.27]	0.096** [4.61]	-0.050** [-24.7]	-0.054** [-2.69]	-0.014 [-0.61]
Year-Qtr Effects	X	X	X	X	X	X	X	X	X
State Effects			X			X			X
N	23,888	23,888	23,013	23,530	23,530	22,664	17,616	17,616	16,831
R ²	0.048	0.053	0.085	0.031	0.033	0.043	0.043	0.043	0.05

	CRE Loans			Deposits Held in Other Banks			Loan Loss Provision Balance		
Post Close	0.031** [5.00]	0.035** [5.59]	0.020* [1.97]	-0.045** [-4.09]	-0.051** [-4.48]	-0.082* [-2.11]	0.048** [3.31]	0.058** [3.74]	0.11* [2.06]
Bank Size		-0.0021 [-1.88]	-0.0046** [-3.47]		0.0032* [2.19]	0.0028 [1.70]		-0.0048** [-3.27]	-0.0069** [-4.04]
Constant	0.016** [12.1]	0.042** [2.90]	0.068** [3.75]	0.070** [37.0]	0.031 [1.70]	0.042 [1.79]	0.062** [35.0]	0.12** [6.51]	0.13** [5.95]
Year-Qtr Effects	X	X	X	X	X	X	X	X	X
State Effects			X			X			X
N	17,978	17,978	17,180	22,998	22,998	22,144	22,942	22,942	22,104
R ²	0.018	0.018	0.027	0.02	0.02	0.021	0.021	0.021	0.028

Notes: These regressions track changes in key bank balance sheet variables around FDIC-assisted merger for the synthetic bank, comprising the problem bank and the acquiring bank, and surviving zombie banks. The left hand side variable in each regression is the log change of the variable at a quarterly frequency. Bank size is the mean of log assets for each bank. For synthetic banks, I include observations for the two years before and after the merger, excluding the period immediately after the merger. Standard errors are clustered at the bank level. T-statistics are displayed in the brackets. The significance levels are: ** p<0.01, * p<0.05.

Synthetic banks reduce their rates of allocation to the safest, most liquid assets, with other bank deposit growth contracting by between 3.4 and 6.5 log points post resolution. This contraction coincides with an aggregate expansion for the universe of healthy banks in deposits held in other banks. Loan loss provisioning continues to increase post resolution. At first blush, this seems inconsistent with most bad loans passing into the receivership and off of the synthetic bank's balance sheet. However, we would expect this pattern if prior to resolution, either zombies were underprovisioning in order to preserve capital or if a higher rate of charge-offs due to defaulting loans counteracted provisioning growth.

3.6.2 Alternative Explanations and Secondary Tests

Without a natural experiment, endogeneity concerns cannot be assumed away and must be addressed directly. I consider three important endogeneity risks. First is the claim that lending would have recovered even without resolution. Second is the related claim that because resolution reduces the size of the loan book, some amount of mechanical mean reversion should be expected. Third is the claim that recovery is driven exclusively by acquirer characteristics and plans.

What is the counterfactual lending path for zombies?

Prior to resolution, the problem loans in zombie bank balance sheets are concentrated in residential and commercial land development and construction loans. These loans are largely problematic because of diminished local opportunities. For this reason, we might expect zombies to be geographically concentrated in areas with limited projects to finance going forward. Relative to other banks, these banks should see the slowest resumption of lending following the Great Slump. Resolution and the new bank's balance sheet clearly alter this course.

A second response pertains to the FDIC's motives. The FDIC's key objective is to protect insured depositors through the lowest cost option from the point of view of the Deposit Insurance Fund (DIF). If opportunities are expected to improve in a bank's area,

the least costly alternative from the FDIC's point of view is forbearance, not resolution and assisted merger. Regulators should only shut down the worst performing banks in the worst performing geographies. So, if there is a selection bias in the sample, it should bias these results against an increase in lending post failure.

Furthermore, the timing of the regulator's decision to close a bank is unlikely to coincide with the exact nadir of lending for the failing bank. Regulators are constrained both by the legislated timeline for dealing with struggling banks and by limited human resources. Once a bank is marked as critically undercapitalized and enters into an action agreement with its regulator, the FDI Act requires the FDIC to begin planning for closure and to close the bank within 270 days, conditional on the agreement not being terminated. In addition, the FDIC tries to group multiple failures in geographically close areas to avoid spreading its human resources too thinly. More fundamentally, the FDIC is less concerned with lending activity than it is with protecting insured deposits and the insurance fund. Thus, observing a collapse in lending would not be sufficient to induce regulatory action.

Can we distinguish mechanical mean reversion from new lending?

Assume that both the zombie and acquirer are properly funding all positive NPV projects prior to resolution. Now, the zombie runs into some bad luck unrelated to the quality of lending opportunities, such as a particularly large individual default or a general deposit run. Regulators might close the zombie and arrange a merger, which would reduce the size of the synthetic loan book since some loans pass into the receivership. Then, the synthetic bank resumes lending at exactly the same gross rate. Because the base loan book is now temporarily smaller, the post resolution rate of growth will look larger even though the bank's set of new projects has not changed.

First, this criticism can not explain a pattern where there is a sign change in activity around the closure date, such as for C&I lending. In this case, the mechanical contraction biases the difference toward zero through increasing the pre failure rate. Second, a comparison of different categories of asset allocation holds constant the contraction in the loan book.

Since some categories contract and some do not grow faster post resolution, my overall findings cannot be a mechanical result of the smaller loan book.

What is the counterfactual lending path for acquirers?

I do not observe the separate activities of the zombie and acquirer banks post merger. More aggressive lending by the acquirer might be sufficient to generate the observed pattern of lending growth. Of course, if this more aggressive lending were due to the new funding to the combined institution that results from resolution, this would be part of the mechanism argued for in this paper. Still, a close reading of my ideal experiment and hypothesis requires that *zombie* lending recover post treatment, holding acquirer characteristics constant.

Table 3.4 presents one attempt to study recovery while holding acquirer characteristics constant. It implements the following thought experiment. If zombie characteristics matter for recovery, then we should see a larger recovery when the zombie is larger relative to the synthetic bank and when the zombie's loan portfolio was worse prior to resolution. I measure zombie size with the ratio of zombie deposits to synthetic bank deposits in the quarter prior to failure. I measure the quality of the zombie's loan portfolio with the ratio of FDIC's estimated losses from the receivership to the size of the zombie's loan book in the quarter prior to failure. The key coefficient of interest is the interaction of zombie size with loss size for the post failure period.

The evidence is broadly consistent with zombie characteristics being important for the post resolution results. A larger zombie bank relative to the synthetic bank coincides with a larger recovery in lending and a larger reduction in lending to other banks. Going from the 25th to the 75th percentile increases the post failure total lending rate by 1.0 ($= 3.6 \times (0.33 - 0.05)$) log points, the C&I lending rate by 1.4 ($= 4.9 \times (0.33 - 0.05)$) log points, and reduces the post failure bank lending rate by 0.3 ($= 1.0 \times (0.33 - 0.05)$) log points. Conversely, the loss size coefficients go in the opposite direction, suggesting that

Table 3.4: Lending and Asset Allocation Post Resolution as a Function of Zombie Characteristics (Within Synthetic Banks)

	Total Loans			C&I Loans			Deposits Held in Other Banks		
Post Close	0.011 [1.31]	0.035** [3.36]	0.0071 [0.75]	0.0082 [0.76]	0.017 [1.16]	-0.0017 [-0.13]	-0.057* [-2.40]	-0.028 [-0.77]	-0.085** [-3.00]
Zombie Size	-0.023* [-2.31]		0.019 [0.76]	-0.034* [-2.47]		-0.015 [-0.60]	0.086* [2.56]		0.11 [1.72]
Post*Zombie Size	0.036 [1.68]			0.049* [2.02]			-0.1 [-1.62]		
Bank Size	-0.00082 [-0.62]	-0.00086 [-0.57]	-0.002 [-1.27]	-0.00064 [-0.39]	-0.00011 [-0.067]	-0.0013 [-0.68]	-0.00054 [-0.15]	-0.0033 [-0.90]	-0.0031 [-0.78]
Loss Size		0.034* [2.18]	0.049** [2.80]		0.024 [1.22]	0.037 [1.70]		0.12** [2.88]	0.091* [2.07]
Post*Loss Size		-0.053* [-1.99]			-0.02 [-0.63]			-0.21** [-2.77]	
Zombie Size*Loss Size			-0.14* [-2.27]			-0.081 [-1.19]			-0.15 [-0.92]
Post*Zombie Size*Loss Size			0.084* [2.03]			0.11* [2.25]			-0.22 [-1.64]
Constant	0.010* [2.56]	-0.0043 [-0.77]	-0.00082 [-0.13]	0.0039 [0.73]	-0.0072 [-1.06]	-0.00091 [-0.12]	0.099 [1.84]	0.13* [2.49]	0.13* [2.16]
Year-Qtr Effects	X	X	X	X	X	X	X	X	X
N	3,038	2,602	2,602	3,038	2,602	2,602	3,070	2,620	2,620
R ²	0.023	0.02	0.024	0.027	0.024	0.026	0.05	0.055	0.055

Notes: These regressions track changes in key bank balance sheet variables around FDIC-assisted merger for the synthetic bank, comprising the problem bank and the acquiring bank. The left hand side variable in each regression is the log change of the variable at a quarterly frequency. Zombie size is the ratio of the zombie bank's deposits to the synthetic bank's deposits in the quarter before closure. Loss size is the ratio of the FDIC's estimated losses from resolution to the zombie's loan total loan book. Bank size is the mean of log assets for each bank. I include observations for the two years before and after the merger, excluding the period immediately after the merger. Standard errors are clustered at the bank level. T-statistics are displayed in the brackets. The significance levels are: ** $p < 0.01$, * $p < 0.05$.

worse performing zombies did significantly more lending prior to failure and show more sluggish lending recovery post failure. However, the interaction coefficients imply that the net effect of these zombie characteristics is to reinforce the recovery. Moving from the 25th to the 75th percentile in the interaction term increase the post failure total lending rate by 1.0 ($= 8.4 \times (0.14 - 0.02)$) log points, the C&I lending rate by 1.32 ($= 11.0 \times (0.14 - 0.02)$) log points, and reduces the post failure bank lending rate by 2.6 ($= 22.0 \times (0.14 - 0.02)$) log points.

An alternative approach to partialing out the expected path for acquirer activity is to use the bidding process to identify a comparable would-be acquirer. The FDIC invites multiple bids during the resolution process in order to minimize the cost to the insurance fund. Subsequent to the merger, the FDIC publishes the list of losing bidder banks and the list of bids. In order to keep the exact bids confidential, the FDIC does not map the bids to the bidders and only publishes the list of bidders when there are two or more alternatives. I use these disclosures to match bidders to actual acquirers. Disclosure restrictions limit my sample to 100 failed banks.

For each failed bank, I generate counterfactual growth rates post failure in the key balance sheet variables by averaging the rates for the bidder banks. Assuming the synthetic growth rates are weighted averages of the zombie and acquirer rates, I can solve for the post failure zombie rate of growth using the formula,

$$r_{Synthetic} = w_{Zombie} \cdot r_{Zombie} + (1 - w_{Zombie}) \cdot r_{Acquirer}, \quad (3.1)$$

where the zombie weight is the zombie size variable described above and the acquirer growth rate is proxied with the average bidder rates. This imputation procedure does not impose restrictions on the sign of the zombie rate relative to the synthetic rate. If bidder banks grow fast enough relative to the synthetic bank, the imputed lending rate for the post failure zombie can be negative.

Table 3.5 presents a panel regression comparing these imputed zombie paths to the path for problem banks. While my other specifications map to experiments in which the synthetic

bank is treated with resolution, this specification measures treatment effects for the sample of zombies. The smaller sample introduces more noise into the estimates, especially in the case with state fixed effects where the number of treated zombies per cell becomes quite small. I present these regressions to preserve consistency.

The results suggest that zombie lending recovers post resolution, even after partialing out a plausible counterfactual lending path for the acquiring bank. Total lending increases by between 5.2 and 15 log points. C&I lending increases between 5.6 and 7.3 log points. The construction lending contraction slows by between 4.9 and 5.8 log points. Though not statistically significant, lending to other banks decreases by between 8.1 and 22 log points.

3.7 Reduced Growth in Zombie-Afflicted Counties

Taken together, the evidence presented so far suggests that by clearing bad loans from zombie balance sheets, the process of FDIC-assisted merger and receivership enables new lending and asset reallocation by zombie banks. Prior to resolution, these banks suffer from an overhang of anticipated losses on past loans and must forgo new lending opportunities as a result. What remains to be shown is whether the phenomenon of loss overhang was pervasive enough to impede economic outcomes during the Great Slump. This conclusion rests on two premises. First, the banking system suffered enough losses to overwhelm the regulatory apparatus for resolving troubled institutions. This allowed a number of banks to continue operating in a stultifying zombie state for months and years. Second, the presence of zombie banks in some counties was large enough to prevent healthy banks from picking up the slack in positive NPV projects. To address the criticism that these were precisely the geographies expected to grow most slowly, I focus on activity in tradable goods industries, which should be responding to macro factors.

Table 3.5: Lending and Asset Allocation Post Resolution (Imputed, Failed Zombies vs. Surviving Zombies)

	Total Loans			C&I Loans			Construction Loans		
Post Close	0.052*	0.055*	0.15	0.056**	0.058**	0.073	0.050**	0.049**	0.058**
	[2.31]	[2.41]	[1.41]	[2.96]	[3.10]	[1.37]	[6.56]	[6.29]	[4.32]
Bank Size		-0.013**	-0.014**		-0.0099**	-0.012**		0.00034	-0.00033
		[-6.85]	[-7.28]		[-5.17]	[-5.71]		[0.21]	[-0.19]
Constant	0.021**	0.17**	0.18**	0.0074**	0.13**	0.13**	-0.050**	-0.054**	-0.014
	[10.1]	[7.33]	[8.24]	[3.37]	[5.23]	[5.19]	[-24.7]	[-2.69]	[-0.61]
Year-Qtr Effects	X	X	X	X	X	X	X	X	X
State Effects			X			X			X
N	21,946	21,946	21,691	21,597	21,597	21,352	17,616	17,616	16,831
R ²	0.054	0.06	0.094	0.034	0.035	0.047	0.043	0.043	0.05

	CRE Loans			Deposits Held in Other Banks			Loan Loss Provision Balance		
Post Close	0.046	0.047	0.11*	-0.081	-0.082	-0.22	0.044	0.045	0.12
	[1.59]	[1.62]	[2.36]	[-1.21]	[-1.22]	[-1.46]	[1.05]	[1.07]	[0.76]
Bank Size		-0.0045**	-0.0061**		0.0038*	0.0040*		-0.0041*	-0.0066**
		[-3.03]	[-3.64]		[2.17]	[2.16]		[-2.03]	[-3.11]
Constant	0.015**	0.069**	0.084**	0.071**	0.026	0.026	0.066**	0.11**	0.13**
	[10.6]	[3.70]	[3.86]	[36.6]	[1.26]	[1.02]	[35.6]	[4.73]	[4.81]
Year-Qtr Effects	X	X	X	X	X	X	X	X	X
State Effects			X			X			X
N	17,050	17,050	16,845	21,877	21,877	21,678	21,900	21,900	21,651
R ²	0.017	0.017	0.029	0.019	0.019	0.021	0.024	0.024	0.031

Notes: These regressions track changes in key bank balance sheet variables around FDIC-assisted merger for failed zombie banks, with post resolution growth imputed using losing bidder data, and surviving zombie banks. The left hand side variable in each regression is the log change of the variable at a quarterly frequency. Bank size is the mean of log assets for each bank. For synthetic banks, I include observations for the two years before and after the merger, excluding the period immediately after the merger. Standard errors are clustered at the bank level. T-statistics are displayed in the brackets. The significance levels are: ** p<0.01, * p<0.05.

3.7.1 FDIC Constraints

When the FDIC closes an institution and initiates a sale, it is obligated to cover the difference between the deposits assumed and assets purchased by the acquirer. Therefore, whether the FDIC has sufficient liquidity to resolve a large number of banks is a first order concern. The FDIC has a Congressionally approved credit line with the US Treasury, which in theory should protect it from a liquidity squeeze. However, FDIC actions during the crisis indicate that they demurred from drawing on this resource.

Historical evidence of FDIC liquidity constraints comes directly from the FDIC's resolution handbook, in its discussion of lessons learned from the savings and loan crisis:

To efficiently resolve a banking crisis, it is critical to have an adequate insurance fund reserve. If such funds are not available, the problems may become worse as a result of delay. As the Federal Savings and Loan Insurance Corporation (FSLIC) began to experience a greater outflow of funds than it had coming in, it took steps to conserve cash. Although many of its programs were designed to give failing institutions time to work out their problems, some programs had the unintended effect of postponing the problems and actually increasing resolution costs. *The FSLIC lacked the financial liquidity to promptly close insolvent institutions, and many of them remained open to compete with healthy institutions.*

During the banking crisis that began in 2007, it is unclear whether the FDIC heeded its own earlier warning. Figure 3.6 shows that the Deposit Insurance Fund (DIF) quickly exhausted its resources as the crisis escalated in 2008 and 2009. Prior to the crisis, the level of the fund balance was approximately \$50 billion or a little over one percent of insured deposits. The deficit that emerged in the third quarter of 2009 owed to a significant accretion of contingent liabilities set aside for future failures. Receivables from resolutions increased as well, but not at a pace sufficient to match the rise in fund liabilities.

To address this deficit, the FDIC was required to outline a restoration plan in the third quarter of 2008. In the fourth quarter of 2009, the FDIC attempted to address its liquidity problem by drawing \$42 billion of future assessments into its account. Arguably, this was not an ideal time for the FDIC to be shoring up the DIF.

More evidence that the resolution pipeline backed up during the crisis is shown in Figure

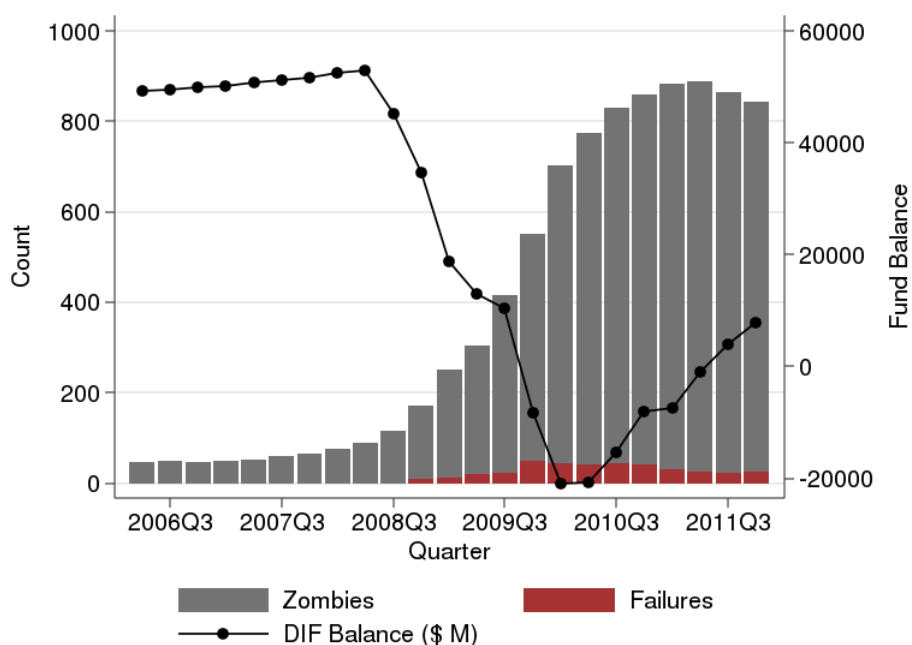


Figure 3.6: *Zombie and Failed Banks against the DIF Balance*

Notes: Data come from the FDIC quarterly CFO reports to the Board of Directors and from the Quarterly Banking Profile.

3.7. I use book equity and risk capital to plot the mean capitalization for zombies just prior to receiving a Prompt Corrective Action notice and just prior to being resolved. The oversight apparatus shows no signs of strain, as capital ratios at the time a bank receives notice remain constant through the crisis. In contrast, it appears that troubled banks remained in trouble for longer as the crisis evolved, possibly due to limited regulatory resources. Early failures have mean capital ratios greater than two percent of assets, while later failures have average capital ratios of just one percent and frequently fail with negative *reported* book equity. Given the likelihood that many of the assets in these failing banks are optimistically marked at the time of failure, these data suggest it was possible for a deeply insolvent bank to continue operating for months and perhaps years without regulators stepping in.

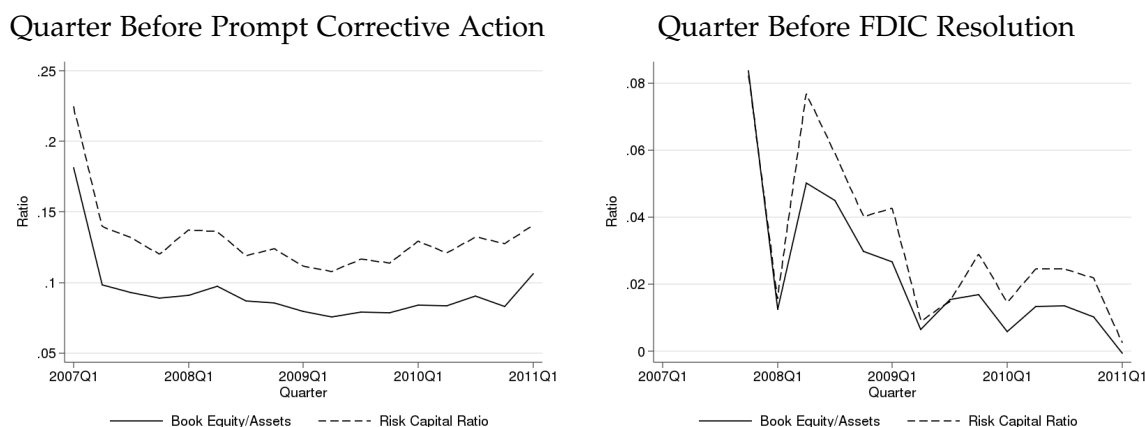


Figure 3.7: *Capital Ratios for Zombie Banks over Time*

Notes: Dates for Prompt Corrective Action come from public disclosures of action notices.

3.7.2 Local Economic Effects of Zombie Banks

Evidence that many zombie banks were allowed to survive despite being unable to extend new loans does not imply that good projects were being ignored. This is because the supply of funds from healthy banks and other capital sources may have been robust enough to absorb the unfunded projects of surviving zombie banks. The surviving zombie phenomenon may have only led to a reallocation of the same projects, perhaps with some destruction of relationship capital, but ultimately with limited real consequences.

How important were zombie banks in their local economies? I use the FDIC's Summary of Deposits to map out the quarterly branch-level incidence of banks operating under regulatory actions. I aggregate these time series to the county level and ask what fraction of local deposits are housed in a zombie bank at a given point in time. Figure 3.8 plots a heat map of zombie deposits as of the fourth quarter of 2010. In the colored counties, at least four percent of bank deposits are held by zombie banks.

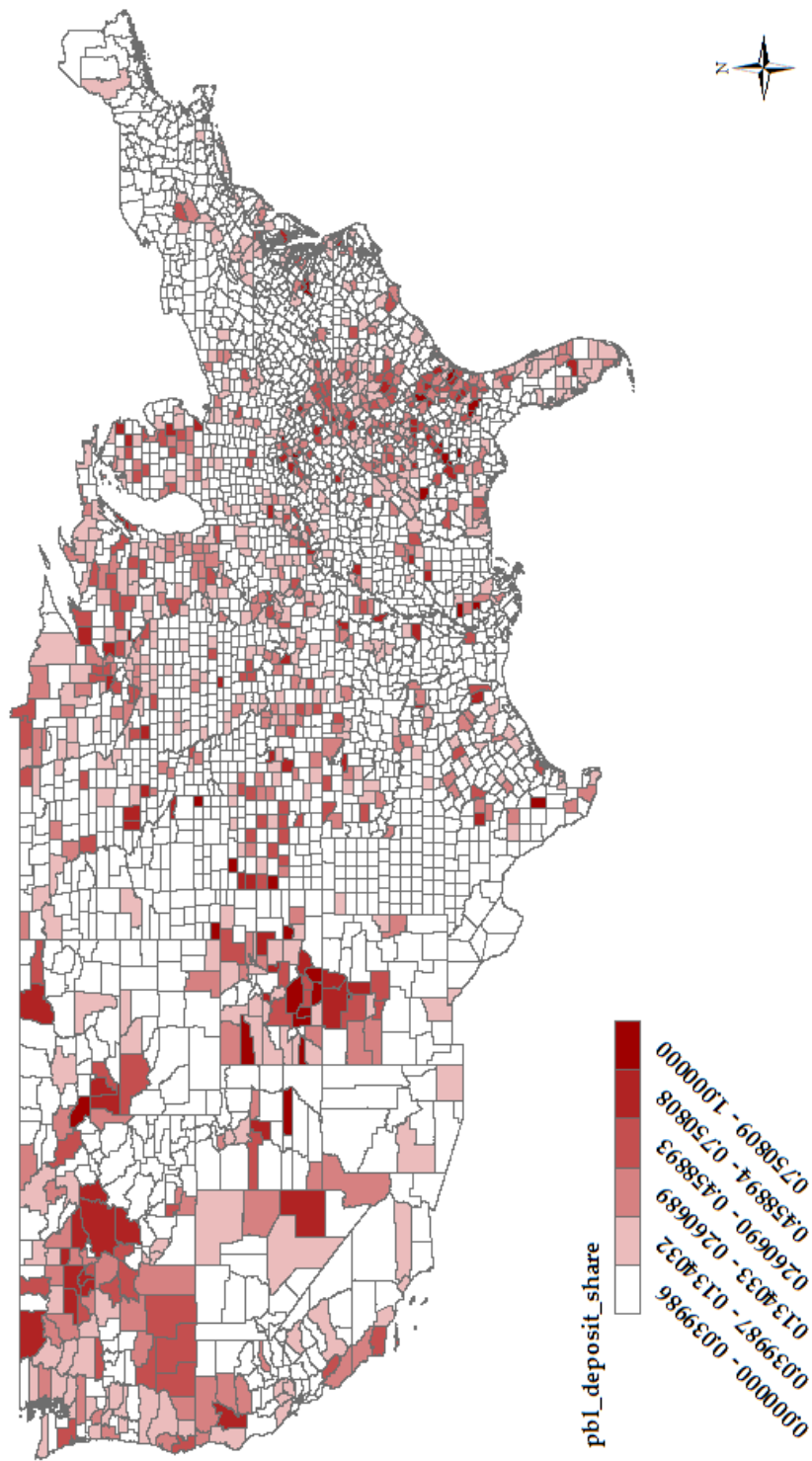


Figure 3.8: Zombie Deposit Shares in U.S. Counties (2010:Q4)

Notes: Branch-level data come from the annual FDIC Summary of Deposits. Zombie banks are banks operating under Prompt Corrective Action or Cease & Desist agreements with their comparison regulators. For each county, the deposit share is the fraction of all deposits in zombie banks.

The map of zombie counties yields two insights. First, the geographic incidence of problem banks is surprisingly spread around the country, with most regions outside the Northeast being marked by zombie banks. Several regions not commonly associated with the housing and subprime booms — the Pacific Northwest, Colorado, Illinois, and Georgia — have many zombie counties. Second, there are many counties in which the concentration of zombie deposits is above ten percent of total area deposits. In these counties, it may be quite possible that the remaining banks are unable to pick up all of the slack left by their zombie competitors. Thus, it is plausible that the persistence of zombie banks has real consequences at the local economic level. The diffuse geographic incidence of these zombies implies that these consequences may even appear in aggregate data.

Table 3.6 presents an attempt at measuring the relationship between the presence of zombie banks and economic outcomes at the county level. I use the establishment survey in the Quarterly Census of Employment and Wages to measure for each county the level of employment disaggregated into sectors. I mark zombie counties using the concentration of county deposits in zombie banks as of the fourth quarter of 2009, which is the first quarter the FDIC's deposit insurance fund turns net worth negative. At this point, with the FDIC's constraints potentially binding, the persistence of surviving zombies is likely to influence subsequent economic activity.

To mitigate the risk of reverse causality, I focus on the mining, services, and manufacturing sectors, which were not directly responsible for the banking problems. In addition, for the mining and manufacturing sectors, national and international demand are likely to influence activity more than local conditions. This helps address concerns about omitted differential demand factors driving the results.

The results in Table 3.6 provide mixed evidence of spillovers from suffering banks into industries not responsible for the crash. In 2010, zombie counties have services employment between 1.7 and 2.5 log points lower than non-zombie counties. Mining employment is

Table 3.6: Employment Paths for Services, Mining, and Manufacturing (Zombie vs. Healthy Counties)

Log(Services Employment)			Log(Mining Employment)			Log(Manufacturing Employment)		
2007Q2	0.050** [30.1]	Z*2007Q2 [-0.0033 [-0.71]	2007Q2	0.11** [30.4]	Z*2007Q2 [0.0096 [0.72]	2007Q2	0.010** [6.11]	Z*2007Q2 [-0.0059 [-0.83]
2007Q3	0.065** [28.8]	Z*2007Q3 [-0.0047 [-0.73]	2007Q3	0.16** [28.3]	Z*2007Q3 [0.018 [0.85]	2007Q3	0.015** [6.15]	Z*2007Q3 [-0.0047 [-0.60]
2007Q4	0.036** [26.8]	Z*2007Q4 [-0.0038 [-0.99]	2007Q4	0.12** [28.9]	Z*2007Q4 [-0.013 [-0.91]	2007Q4	-0.002 [0.80]	Z*2007Q4 [-0.0038 [-0.53]
2008Q1	0.0056** [4.04]	Z*2008Q1 [-0.0038 [-0.92]	2008Q1	0.021** [5.21]	Z*2008Q1 [-0.01 [-0.65]	2008Q1	-0.023** [-7.66]	Z*2008Q1 [0.01 [1.00]
2008Q2	0.048** [22.3]	Z*2008Q2 [-0.0083 [-1.32]	2008Q2	0.13** [24.2]	Z*2008Q2 [-0.0056 [-0.28]	2008Q2	-0.019** [-5.59]	Z*2008Q2 [0.0083 [0.69]
2008Q3	0.059** [22.5]	Z*2008Q3 [-0.0084 [-1.11]	2008Q3	0.17** [25.3]	Z*2008Q3 [0.009 [0.38]	2008Q3	-0.025** [-6.46]	Z*2008Q3 [0.01 [0.77]
2008Q4	0.020** [10.7]	Z*2008Q4 [-0.011 [-1.92]	2008Q4	0.13** [22.4]	Z*2008Q4 [-0.017 [-0.82]	2008Q4	-0.063** [-15.7]	Z*2008Q4 [0.012 [0.92]
2009Q1	-0.025** [-14.1]	Z*2009Q1 [-0.013* [-2.21]	2009Q1	-0.0024 [-0.42]	Z*2009Q1 [-0.039 [-1.83]	2009Q1	-0.14** [-30.1]	Z*2009Q1 [0.0079 [0.54]
2009Q2	0.014** [5.90]	Z*2009Q2 [-0.018* [-2.38]	2009Q2	0.083** [12.9]	Z*2009Q2 [-0.025 [-1.03]	2009Q2	-0.17** [-34.6]	Z*2009Q2 [0.012 [0.79]
2009Q3	0.024** [8.83]	Z*2009Q3 [-0.013 [-1.49]	2009Q3	0.11** [14.8]	Z*2009Q3 [-0.004 [-1.14]	2009Q3	-0.19** [-33.9]	Z*2009Q3 [0.012 [0.71]
2009Q4	-0.014** [-6.46]	Z*2009Q4 [-0.015* [-2.06]	2009Q4	0.085** [12.2]	Z*2009Q4 [-0.042 [-1.66]	2009Q4	-0.20** [-37.5]	Z*2009Q4 [0.0087 [0.52]
2010Q1	-0.049** [-22.3]	Z*2010Q1 [-0.017** [-2.64]	2010Q1	-0.027** [-3.96]	Z*2010Q1 [-0.04 [-1.64]	2010Q1	-0.23** [-39.3]	Z*2010Q1 [-0.012 [-0.64]
2010Q2	0.0042 [1.64]	Z*2010Q2 [-0.022** [-2.74]	2010Q2	0.11** [14.3]	Z*2010Q2 [-0.036 [-1.34]	2010Q2	-0.20** [-34.6]	Z*2010Q2 [-0.014 [-0.75]
2010Q3	0.020** [6.83]	Z*2010Q3 [-0.021* [-2.32]	2010Q3	0.15** [18.2]	Z*2010Q3 [-0.035 [-1.13]	2010Q3	-0.19** [-31.1]	Z*2010Q3 [-0.019 [-0.97]
2010Q4	-0.0077** [-3.02]	Z*2010Q4 [-0.025** [-3.23]	2010Q4	0.12** [15.3]	Z*2010Q4 [-0.059* [-2.13]	2010Q4	-0.19** [-32.7]	Z*2010Q4 [-0.016 [-0.82]
Constant			Constant			Constant		
State Effects		X	State Effects		X	State Effects		X
N		49,742	N		44,914	N		44,475
R ²		0.15	R ²		0.083	R ²		0.309

Notes: These regressions track changes in employment in the services, resource and mining, and manufacturing sectors at the county level. The left hand side variable in each regression is the log level of employment at a quarterly frequency from 2007 through 2010. The interaction terms mark zombie counties as counties with top decile fractions of deposits in zombie banks (zombie share > 13 percent) as of 2009:Q4. Standard errors are clustered at the county level. T-statistics are displayed in the brackets. The significance levels are: ** $p < 0.01$, * $p < 0.05$.

between 3.5 and 5.9 log points lower, and manufacturing employment is between 1.2 and 1.9 log points lower. However, the results for manufacturing and mining are not statistically distinguishable from non-zombie counties.

3.8 Conclusion

This paper addresses a prominent puzzle in macroeconomics and finance: why are economies slow to recover from recessions that coincide with financial crises? It offers evidence consistent with the idea that loss overhang in the banking sector is an important explanation. Zombie banks impaired by unrecognized losses restrict new lending in healthy categories, prop up lending in unhealthy categories, and overallocate to safe, liquid assets. FDIC-induced failure reverses this pattern by carving out bad loans and trading them for cash. This allows lending to recover.

The second part of the paper shows that regulatory constraints, in particular the FDIC's limited deposit insurance fund, slowed the recovery and prolonged the Great Slump. These constraints led to slower employment and establishment growth in counties afflicted with unhealed zombies. These results underscore the importance of ensuring that regulators have sufficient capacity during a crisis to execute their duties.

References

- Abel, Andrew B.** 1982. "Dynamic Effects of Permanent and Temporary Tax Policies in a q Model of Investment." *Journal of Monetary Economics*, 9: 353–373.
- Abel, Andrew B.** 1990. "Consumption and Investment." *Handbook of Monetary Economics*.
- Abel, Andrew B., and Janice C. Eberly.** 1994. "A Unified Model of Investment Under Uncertainty." *American Economic Review*, 1369–1384.
- Abreu, Dilip, and Markus K. Brunnermeier.** 2003. "Bubbles and Crashes." *Econometrica*, 71(1): 173–204.
- Akerlof, George A.** 1991. "Procrastination and obedience." *American Economic Review*, 81(2): 1–19.
- Allen, Franklin, Stephen Morris, and Hyun Song Shin.** 2006. "Beauty Contests and Iterated Expectations in Asset Markets." *Review of Financial Studies*, 19(3): 719–752.
- Almeida, Heitor, Murillo Campello, and Michael S. Weisbach.** 2004. "The Cash Flow Sensitivity of Cash." *Journal of Finance*, 59(4): 1777–1804.
- Ashcraft, Adam.** 2005. "Are banks really special? New evidence from the FDIC-induced failure of healthy banks." *American Economic Review*, 95(5): 1712–1730.
- Auerbach, Alan J., and James M. Poterba.** 1987. "Tax loss carryforwards and corporate tax incentives." In *The effects of taxation on capital accumulation*. 305–342. University of Chicago Press.
- Auerbach, Alan J., and Kevin A. Hassett.** 1992. "Tax policy and business fixed investment in the United States." *Journal of Public Economics*, 47(2): 141–170.
- Bayer, Patrick, Christopher Geissler, and James W. Roberts.** 2013. "Speculators and Middlemen: The Role of Intermediaries in the Housing Market." *Economics Research Initiatives at Duke Working Paper* 93.
- Bernanke, Ben.** 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review*, 73(3): 257–276.
- Bernanke, Ben, and Alan Blinder.** 1992. "The federal funds rate and the channels of monetary policy." *American Economic Review*, 82(4): 901–921.

- Bernanke, Ben, and Mark Gertler.** 1989. "Agency costs, net worth, and business fluctuations." *American Economic Review*, 79(1): 14–31.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan.** 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249–275.
- Bond, Stephen, and John Van Reenen.** 2007. "Microeconomic Models of Investment and Employment." In *Handbook of Econometrics*. Vol. 6, Chapter 65, 4417–4498.
- Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner.** 2006. "Causes of Sprawl: A Portrait from Space." *Quarterly Journal of Economics*, 121(2): 587–633.
- Burnside, Craig, Martin Eichenbaum, and Sergio Rebelo.** 2013. "Understanding Booms and Busts in Housing Markets." *National Bureau of Economic Research Working Paper* 16734.
- Caballero, Ricardo J., and Eduardo MRA Engel.** 1999. "Explaining investment dynamics in US manufacturing: a generalized (S, s) approach." *Econometrica*, 67(4): 783–826.
- Caballero, Ricardo, Takeo Hoshi, and Anil Kashyap.** 2008. "Zombie Lending and Depressed Restructuring in Japan." *American Economic Review*, 98(5): 1943–1977.
- Case, Karl E., Robert J. Shiller, and Anne K. Thompson.** 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity*, 265–298.
- Chaney, Thomas, David Sraer, and David Thesmar.** 2012. "The Collateral Channel: How Real Estate Shocks affect Corporate Investment." *American Economic Review*, 102(6): 2381–2409.
- Cheng, Ing-Haw, Sahil Raina, and Wei Xiong.** 2014. "Wall Street and the Housing Bubble." *American Economic Review*, Forthcoming.
- Chen, Joseph, Harrison Hong, and Jeremy C. Stein.** 2002. "Breadth of Ownership and Stock Returns." *Journal of Financial Economics*, 66: 171–205.
- Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145–1177.
- Chetty, Raj, John Friedman, and Emmanuel Saez.** 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review*.
- Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri.** 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics*, 126(2): 749–804.
- Chirinko, Robert S., Steven M. Fazzari, and Andrew P. Meyer.** 1999. "How responsive is business capital formation to its user cost? An exploration with micro data." *Journal of Public Economics*, 74: 53–80.

- Committee on Ways & Means.** 2003. "Report 108-94: Jobs and Growth Reconciliation Tax Act of 2003." *U.S. House of Representatives*.
- Cooper, Russell, and John Haltiwanger.** 2006. "On the Nature of Capital Adjustment Costs." *Review of Economic Studies*, 73(3): 611–633.
- Cummins, Jason G., Kevin A. Hassett, and R. Glenn Hubbard.** 1994. "A Reconsideration of Investment Behavior Using Tax Reforms as Natural Experiments." *Brookings Papers on Economic Activity*, 2: 1–74.
- Cummins, Jason G., Kevin A. Hassett, and R. Glenn Hubbard.** 1995. "Have tax reforms affected investment?" In *Tax Policy and the Economy*, Volume 9. 131–150. MIT Press.
- Cummins, Jason G., Kevin A. Hassett, and R. Glenn Hubbard.** 1996. "Tax Reforms and Investment: A Cross-Country Comparison." *Journal of Public Economics*, 62(1-2): 237–273.
- Davis, Morris A., and Jonathan Heathcote.** 2007. "The Price and Quantity of Residential Land in the United States." *Journal of Monetary Economics*, 54(8): 2595–2620.
- D'Avolio, Gene.** 2002. "The Market for Borrowing Stock." *Journal of Financial Economics*, 66(2): 271–306.
- DeFusco, Anthony, Wenjie Ding, Fernando Ferreira, and Joseph Gyourko.** 2013. "The Role of Contagion in the Last American Housing Cycle." *Unpublished*.
- Desai, Mihir A., and Austan D. Goolsbee.** 2004. "Investment, overhang, and tax policy." *Brookings Papers on Economic Activity*, 35(2): 285–355.
- Dew-Becker, Ian.** 2012. "Investment and the Cost of Capital in the Cross-Section: The Term Spread Predicts the Duration of Investment." *mimeo*.
- Edgerton, Jesse.** 2010. "Investment incentives and corporate tax asymmetries." *Journal of Public Economics*, 94(11-12): 936–952.
- Fama, Eugene F.** 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, 25(2): 383–417.
- Farre-Mensa, Joan, and Alexander Ljungqvist.** 2013. "Do Measures of Financial Constraints Measure Financial Constraints?" *NBER Working Paper No. 19551*.
- Faulkender, Michael, and Rong Wang.** 2006. "Corporate financial policy and the value of cash." *Journal of Finance*, 61(4): 1957–1990.
- Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen.** 1988a. "Financing Constraints and Corporate Investment." *Brookings Papers on Economic Activity*, 1988(1): 141–195.
- Fazzari, Steven M., R. Glenn Hubbard, and Bruce C. Petersen.** 1988b. "Investment, Financing Decisions, and Tax Policy." *American Economic Review: Papers and Proceedings*, 78(2): 200–205.
- Gao, Zhenyu.** 2014. "Housing Boom and Bust with Elastic Supplies." *Unpublished*.

- Geanakoplos, John.** 2009. "The Leverage Cycle." *NBER Macroeconomic Annual*, 1–65.
- Glaeser, Edward L.** 2013. "A Nation of Gamblers: Real Estate Speculation and American History." *American Economic Review*, 103(3): 1–42.
- Glaeser, Edward L., and Janet E. Kohlhase.** 2004. "Cities, Regions, and the Decline of Transport Costs." *Papers in Regional Science*, 83(1): 197–228.
- Glaeser, Edward L., and Joseph Gyourko.** 2009. "Arbitrage in Housing Markets." In *Housing Markets and the Economy: Risk, Regulation, and Policy*. Chapter 5, 113–146. Lincoln Institute of Land Policy.
- Glaeser, Edward L., and Matthew E. Kahn.** 2004. "Sprawl and Urban Growth." In *Handbook of Regional and Urban Economics*. Vol. 4, Chapter 56, 2481–2527. North Holland.
- Glaeser, E L, and J Gyourko.** 2005. "Urban Decline and Durable Housing." *Journal of Political Economy*, 113(2): 31.
- Glaeser, E L, J Gyourko, and A Saiz.** 2008. "Housing supply and housing bubbles." *Journal of Urban Economics*, 64(2): 198–217.
- Goolsbee, Austan.** 1998. "Investment Tax Incentives, Prices, and the Supply of Capital Goods." *Quarterly Journal of Economics*, 113(1): 121–148.
- Graham, John R.** 1996. "Proxies for the corporate marginal tax rate." *Journal of Financial Economics*, 42(2): 187–221.
- Graham, John R.** 2000. "How big are the tax benefits of debt?" *Journal of Finance*, 55(5): 1901–1941.
- Grenadier, Steven R.** 1996. "The Strategic Exercise of Options: Development Cascades and Overbuilding in Real Estate Markets." *Journal of Finance*, 51(5): 1653–1679.
- Grossman, S.** 1976. "On the efficiency of competitive stock markets where trades have diverse information." *Journal of Finance*, 31(2): 573–585.
- Guerrieri, Veronica, Daniel Hartley, and Erik Hurst.** 2013. "Endogenous Gentrification and Housing Price Dynamics." *Journal of Public Economics*, 100: 45–60.
- Gyourko, Joseph.** 2009. "Housing Supply." *Annual Review of Economics*, 1: 295–318.
- Gyourko, Joseph, and Albert Saiz.** 2006. "Construction Costs and the Supply of Housing Structure." *Journal of Regional Science*, 46(4): 661–680.
- Hall, Robert.** 2011. "The Long Slump." *American Economic Review*, 101(April): 431–469.
- Hall, Robert E., and Dale W. Jorgenson.** 1967. "Tax Policy and Investment Behavior." *American Economic Review*, 57(3): 391–414.
- Harrison, J M, and D M Kreps.** 1978. "Speculative investor behavior in a stock market with heterogeneous expectations." *Quarterly Journal of Economics*, 92(2): 323.

- Hassett, Kevin A., and R. Glenn Hubbard.** 2002. "Tax Policy and Business Investment." *Handbook of Public Economics*, 3.
- Hausman, Jerry A.** 1981. "Labor Supply." In *How Taxes Affect Economic Behavior.*, ed. Henry Aaron and Joseph Pechman, 27–71. Washington, D.C.:Brookings Institution Press.
- Hayashi, Fumio.** 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica*, 50(1): 213–224.
- Hellwig, Martin F.** 1980. "On the Aggregation of Information in Competitive Markets." *Journal of Economic Theory*, 22(3): 477–498.
- Henderson, J. Vernon, and Yannis M. Ioannides.** 1983. "A Model of Housing Tenure Choice." *American Economic Review*, 73(1): 98–113.
- Hoberg, Gerard, and Gordon Phillips.** 2010. "Real and Financial Industry Booms and Busts." *Journal of Finance*, 65(1): 45–86.
- Hong, Harrison, and David Sraer.** 2012. "Speculative Betas." *National Bureau of Economic Research Working Paper 18548*.
- Hong, H, J Scheinkman, and W Xiong.** 2006. "Asset float and speculative bubbles." *The journal of finance*, 61(3): 1073–1117.
- Hoshi, Takeo, and Anil Kashyap.** 2010. "Will the US bank recapitalization succeed? Eight lessons from Japan." *Journal of Financial Economics*, 97(3): 398–417.
- House, Christopher, and Matthew Shapiro.** 2008. "Temporary Investment Tax Incentives: Theory with Evidence from Bonus Depreciation." *American Economic Review*, 98(3): 737–68.
- Kaplan, Steven N., and Luigi Zingales.** 1997. "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints?" *Quarterly Journal of Economics*, 112(1): 169–215.
- Kashyap, Anil, and Jeremy Stein.** 2000. "What Do a Million Observations on Banks Say About the Transmission of Monetary Policy?" *American Economic Review*, 90(3): 398–417.
- Kitchen, John, and Matthew Knittel.** 2011. "Business Use of Special Provisions for Accelerated Depreciation: Section 179 Expensing and Bonus Depreciation, 2002-2009." *Working paper*.
- Knittel, Matthew, Susan Nelson, Jason Debacker, John Kitchen, James Pearce, and Richard Prisinzano.** 2011. "Methodology to Identify Small Businesses and Their Owners." *OTA Working Paper*, , (August).
- Koijen, Ralph SJ, and Motohiro Yogo.** 2012. "The cost of financial frictions for life insurers." *NBER Working Paper No. 18321*.
- Laibson, David I.** 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2): 443–477.

- Laibson D. Repetto, A, and J Tobacman.** 2007. "Estimating Discount Functions with Consumption Choices over the Lifecycle." *Economics Series Working Papers*.
- Lamont, Owen.** 1997. "Cash Flow and Investment: Evidence from Internal Capital Markets." *Journal of Finance*, 52(1): 83–109.
- Land Advisors.** 2010. "Metro Phoenix Market Overview: First Quarter 2010." Land Advisors Organization.
- Landvoigt, Tim.** 2011. "Housing Demand during the Boom: The Role of Expectations and Credit Constraints." *Unpublished*.
- Landvoigt, Tim, Monika Piazzesi, and Martin Schneider.** 2013. "The Housing Market(s) of San Diego." *National Bureau of Economic Research Working Paper 17723*.
- Leamer, E.** 2007. "Housing is the business cycle." *NBER Working Paper*, 13428.
- Mayer, Christopher J., and C. Tsurriel Somerville.** 2000. "Land Use Regulation and New Construction." *Regional Science and Urban Economics*, 30(6): 639–662.
- McKinley, Jesse, and Griffin Palmer.** 2007. "Nevada Learns to Cash In on Sales of Federal Land." *New York Times*.
- Mian, Atif, and Amir Sufi.** 2009. "The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis." *Quarterly Journal of Economics*, 124(4): 1449.
- Mian, Atif, and Amir Sufi.** 2011. "House Prices, Home Equity-Based Borrowing, and the U.S. Household Leverage Crisis." *American Economic Review*, 101(5): 2132–56.
- Mian, Atif, and Amir Sufi.** 2012. "What Explains High Unemployment? The Aggregate Demand Channel." *National Bureau of Economic Research Working Paper 17830*.
- Mian, Atif, Kamalesh Rao, and Amir Sufi.** 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics*, 128(4): 1687–1726.
- Miller, E M.** 1977. "Risk, uncertainty, and divergence of opinion." *Journal of Finance*, 32(4): 1151–1168.
- Mills, Lillian, Kaye Newberry, and William Trautman.** 2002. "Trends in Book-Tax Income and Balance Sheet Differences." *Tax Notes*, 96(8).
- Morris, S.** 1996. "Speculative investor behavior and learning." *Quarterly Journal of Economics*, 111(4): 1111.
- Myers, Stewart.** 1977. "Determinants of corporate borrowing." *Journal of Financial Economics*, 5(2): 147–175.
- Nichols, Joseph B., Stephen D. Oliner, and Michael R. Mulhall.** 2013. "Swings in Commercial and Residential Land Prices in the United States." *Journal of Urban Economics*, 73(1): 57–76.

- Onsted, Jeffrey A.** 2009. *The Effectiveness of the Williamson Act: A Spatial Analysis*. VDM Verlag Dr. Müller.
- Pástor, Ľuboš, and Pietro Veronesi.** 2003. "Stock Valuation and Learning about Profitability." *Journal of Finance*, 58(5): 1749–1790.
- Pástor, Ľuboš, and Pietro Veronesi.** 2009. "Technological Revolutions and Stock Prices." *American Economic Review*, 99(4): 1451–1483.
- Peek, Joe, and Eric Rosengren.** 2000. "Collateral Damage: Effects of the Japanese Bank Crisis on Real Activity in the United States." *American Economic Review*, 30–45.
- Piazzesi, Monika, and Martin Schneider.** 2009. "Momentum Traders in the Housing Market: Survey Evidence and a Search Model." *American Economic Review*, 99(2): 406–411.
- Poterba, J M.** 1984. "Tax subsidies to owner-occupied housing: an asset-market approach." *Quarterly Journal of Economics*, 99(4): 729.
- Rauh, Joshua D.** 2006. "Investment and financing constraints: Evidence from the funding of corporate pension plans." *Journal of Finance*, 61(1): 33–72.
- Reinhart, Carmen, and Kenneth Rogoff.** 2009. *This time is different: Eight centuries of financial folly*. Princeton University Press.
- RTCSNV.** 2012. "Regional Transportation Plan 2013-2035." Regional Transportation Commission of Southern Nevada.
- Saez, Emmanuel.** 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2(3): 180–212.
- Saiz, A.** 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics*, 125(3): 1253–1296.
- Scheinkman, J A, and W Xiong.** 2003. "Overconfidence and speculative bubbles." *Journal of Political Economy*, 111(6): 1183–1220.
- Segel, Arthur I., Gregory S. Feldman, James T. Liu, and Elizabeth C. Williamson.** 2011. *Stuyvesant Town - Peter Cooper Village: America's Largest Foreclosure*. Harvard Business School.
- Shiller, Robert J.** 2005. *Irrational Exuberance*. Princeton University Press.
- Simsek, Alp.** 2013a. "Belief Disagreements and Collateral Constraints." *Econometrica*, 81(1): 1–53.
- Simsek, Alp.** 2013b. "Speculation and Risk-Sharing with New Financial Assets." *Quarterly Journal of Economics*, 128(3): 1365–1396.
- Somerville, C. Tsurriel.** 1999. "The Industrial Organization of Housing Supply: Market Activity, Land Supply and the Size of Homebuilder Firms." *Real Estate Economics*, 27(4): 669–694.

- Soo, Cindy K.** 2013. "Quantifying Animal Spirits: News Media and Sentiment in the Housing Market." *Unpublished*.
- Stein, Jeremy C.** 1989. "Efficient capital markets, inefficient firms: A model of myopic corporate behavior." *Quarterly Journal of Economics*, 104(4): 655–669.
- Stein, Jeremy C.** 2003. "Agency, information and corporate investment." *Handbook of the Economics of Finance*, 1: 109–163.
- Suher, Michael.** 2013. "Future House Price Expectations in the Recent Boom and Bust." *Unpublished*.
- Summers, Lawrence H.** 1981. "Taxation and Corporate Investment: A q-Theory Approach." *Brookings Papers on Economic Activity*, 1981(1): 67–140.
- Summers, Lawrence H.** 1987. "Investment incentives and the discounting of depreciation allowances." In *The effects of taxation on capital accumulation*. 295–304. University of Chicago Press.
- Titman, Sheridan.** 1983. "Urban Land Prices Under Uncertainty." *American Economic Review*, 75(3): 505–514.
- Tobin, James.** 1969. "A General Equilibrium Approach to Monetary Theory." *Journal of Money, Credit and Banking* *Credit and Banking*, 1(1): 15–29.
- Yagan, Danny.** 2013. "Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut." *mimeo*.

Appendix A

Appendix to Chapter 1

A.1 Investment with Adjustment Costs and a Borrowing Constraint

We develop an infinite horizon, non-stochastic investment model, deriving the testable hypotheses in Section 1.2 and the empirical moments for calibration in Section 1.5.3. The model nests the standard neoclassical investment model with adjustment costs (Hayashi, 1982), a model with credit constraints and a model with managerial myopia.

A.1.1 General Setup

We begin with a discrete time version of Hayashi (1982). Firm value, V_0 , is given by an infinite series of discounted net receipts, R_t . The discount rate, r_t , is risk-adjusted and possibly time varying. The expression for firm value is

$$V_0 = \sum_{t=0}^{\infty} \frac{1}{\prod_{s=0}^t (1 + r_s)} R_t. \quad (\text{A.1})$$

Net receipts in each period reflect net revenues after taxes, investment costs, adjustment costs and depreciation deductions for current and past investments:

$$R_t = [1 - \tau_t] \pi_t - [1 - k_t] p_{I,t} I_t - \psi_t(I_t, K_t) + \tau_t \sum_{x=0}^{\infty} D_{t-x}(x) p_{I,t-x} I_{t-x}, \quad (\text{A.2})$$

where τ_t is the corporate tax rate, π_t is pretax profits, $p_{I,t}$ is the price of investment goods, k_t is the investment tax credit, I_t is investment, ψ_t is adjustment costs and $D_{t-x}(x)$ is the depreciation deduction for capital of age x , based on the schedule from time $t - x$. Pretax profits are π_t , which equals gross revenues, $p_t F_t(K_t, N_t)$, with capital, K_t , and labor, N_t , inputs, less the cost of labor inputs. Net revenues are thus given by

$$\pi_t = p_t F_t(K_t, N_t) - w_t N_t. \quad (\text{A.3})$$

Firms are price takers so output prices, p_t , and wages, w_t , are exogenous. F_t is weakly concave. The firm maximizes (A.1) subject to a capital accumulation law of motion:

$$K_{t+1} = K_t - \delta K_t + I_t, \quad (\text{A.4})$$

where δ is the rate of economic depreciation. The adjustment cost function is convex and reflects after-tax resource losses due to production disruptions and installation.¹

It is useful to have an expression for the stream of future depreciation deductions owed for investment in time t :

$$z_t^0(\beta) = \tau_t D_0 + \beta \sum_{x=1}^{\infty} \frac{1}{\prod_{s=1}^x (1 + r_{t+s})} \tau_{t+x} D_t(x). \quad (\text{A.5})$$

$z_t^0(\beta)$ reflects the present discounted value of one dollar of investment deductions after tax.² If the firm can immediately deduct the full dollar, then z_t^0 equals τ_t . In general, the stream of future deductions will depend on future tax rates and interest rates. β is an additional discount term between zero and one, which reflects the possibility of myopia. We use our heterogeneity analysis to identify this term separately.

¹Hayashi (1982) models adjustment costs through influencing the law of motion in (A.4), rather than as a net receipts flow. Abel (1982) models adjustment costs through augmenting pretax profits in (A.3). There is no strong a priori argument for one versus the other. We adopt this notation to simplify the borrowing constraint in our calibration exercise. Intuitively, it means adjustment costs are not verifiable and thus the firm cannot borrow to offset them. It makes sense to further assume that such costs would not be deductible as well. The hypotheses we derive do not depend on the assumption.

²In the main text, we define z without incorporating the tax rates, in order to isolate the direct effect of bonus. Here, we define z with tax rates because it matches Hayashi (1982)'s notation and highlights the general dependence of the term on future tax rates.

Bonus depreciation, the policy we study in our empirical analysis, allows the firm to deduct a per dollar bonus, θ_t , at the time of the investment and then depreciate the remaining $1 - \theta_t$ according to the normal schedule:

$$z_t(\beta) = \tau_t \theta_t + (1 - \theta_t) z_t^0(\beta) \quad (\text{A.6})$$

At different points in time, Congress set θ_t equal to 0.3, 0.5 or 1. We use these policy shocks to identify the effect of bonus depreciation on investment.

We further generalize z_t by incorporating a nontaxable state. When the next dollar of investment does not affect this year's tax bill, then the firm must carry forward the deductions to future years. Our general z_t reflects this case:

$$z_t(\beta, \gamma) = \gamma z_t(\beta) + (1 - \gamma) \beta \phi z_t(1), \quad (\text{A.7})$$

where $\gamma \in \{0, 1\}$ is an indicator for current tax state and ϕ is a discount factor that reflects both the expected arrival time and the discount rate, r_T , applied to the future period when the firm switches. Note that for the nontaxable firm, β will apply to all future deductions.³

Hayashi (1982) considers the case with β and γ equal to one. We consider this case first. Define $z_t \equiv z_t(1, 1)$. We can rewrite the objective in (A.1) as

$$V_0 = \sum_{t=0}^{\infty} \frac{1}{\prod_{s=0}^t (1 + r_s)} [(1 - \tau_t) \pi_t - \psi_t(I_t, K_t) - (1 - k_t - z_t) p_{I,t} I_t] + A_0, \quad (\text{A.8})$$

where A_0 is the present value of depreciation deductions on past investments.⁴ We assume r is fixed over time and that k equals zero, since the investment tax credit is not active during our sample frame. We isolate the terms where period t investment enters and rewrite the

³This formula is not exactly correct because additional periods will lead to additional accumulated losses for subsequent deductions. The firm will deduct these at an accelerated rate relative to the schedule in $z_t(1)$. This formulation simplifies the algebra and biases our empirical findings toward the neoclassical benchmark.

⁴The A_0 term is important for Hayashi (1982) because it influences the average value of the firm and one purpose of his study is to show when average Q and marginal q are equal. A_0 does not affect the investment decision problem.

relevant part of the problem:

$$\max_I \left\{ -\psi(I, K) - (1 - z)p_I I + \frac{q_{t+1}I}{1 + r} \right\}, \quad (\text{A.9})$$

where q_{t+1} is the multiplier on the law of motion for capital.

We write the first order condition for investment as

$$q_{t+1} = (1 + r) [\psi_I + (1 - z)p_I], \quad (\text{A.10})$$

which emphasizes that optimal investment equates the marginal product of capital, q_{t+1} , with the hurdle rate $(1 + r)$ applied to the marginal costs of investment. These costs include adjustment costs and the price of investment less the value of investment as a tax shield. q_{t+1} is the marginal value of a unit of capital, which accumulates over many future periods. We can apply the envelope condition and differentiate $V_0(K_t) = \max_I V_0(K_t, I)$ to show that

$$q_t = \sum_{s=t}^{\infty} \frac{1}{\Pi_{v=t}^s (1 + r_v + \delta)} [(1 - \tau_s)\pi_{K,s} - \psi_{K,s}], \quad (\text{A.11})$$

which says that q_t includes the present discounted value of future after-tax marginal products for capital, accounting for the rate of economic depreciation.⁵ In a two period model without adjustment costs, we could rewrite (A.10) as

$$r = \left(\frac{1 - \tau}{1 - z} \right) \frac{\pi_{K,t+1}}{p_I} - 1, \quad (\text{A.12})$$

which shows that the general condition is just a dynamic statement of the simple idea that optimal investment should equate returns and the risk-adjusted discount rate.⁶

We augment the problem to introduce the possibility of imperfect capital markets, which leads to a generalized version of (A.10). Firms face a credit limit on gross borrowing, B_t , which accumulates according to

$$B_{t+1} = B_t + (1 - \tau_t)\pi_t - (1 - z_t)p_{I,t}I_t. \quad (\text{A.13})$$

⁵Note that capital also has an effect on future adjustment costs.

⁶Also, note that with immediate expensing, $z = \tau$ and so taxes do not affect investment. This also holds in certain versions of the more general model. See Abel (1982).

Firms must borrow to cover tax obligations and investment outlays, to the extent these exceed current cash flows. Note that z_t and not just $\tau\theta_t$ enters here. This is because future borrowing constraints also matter.

From Summers (1981) to Edgerton (2010), modern empirical studies of investment apply a parameterized version of (A.10), typically under the conditions shown in Hayashi (1982) to yield marginal q equal to average Q .⁷ The financial constraint augmented first order condition is

$$q_{t+1} = (1 + r) [\psi_I + (1 + \lambda)(1 - z)p_I], \quad (\text{A.14})$$

where $\lambda \geq 0$ is the shadow price associated with the borrowing constraint (A.13).⁸ The shadow price on the borrowing constraint works in this model much like a discount rate. To see this, note that without adjustment costs and in the one shot model we can rewrite (A.12) as

$$r + \lambda = \left(\frac{1 - \tau}{1 - z} \right) \frac{\pi_{K,t+1}}{p_I} - 1, \quad (\text{A.15})$$

where we have assumed for illustration that $r\lambda$ is small. The hurdle rate for an investment project reflects both the discount rate and the borrowing spread. In our empirical analysis, we assume that firms use the same r but may differ in λ , in order to back out an implied λ spread between constrained and unconstrained firms.⁹

A.1.2 Testable Hypotheses

We can derive the three testable hypotheses outlined in Section 1.2. Each hypothesis results from defining optimal investment in (A.14) as a function of an exogenous parameter, a , and then implicitly differentiating. The general condition is

$$\psi_{II} \frac{\partial I}{\partial a} + \frac{\partial q}{\partial a} = (1 + \lambda)p_I \frac{\partial z}{\partial a}, \quad (\text{A.16})$$

⁷These assumptions include making firms price takers in all markets and linear homogeneity for production (i.e., constant returns to scale) and adjustment costs.

⁸The general version of (A.9) is $\max_I \left\{ -\psi(I, K) - (1 - z)p_I I + \frac{q_{t+1}I}{1+r} - \lambda(1 - z)p_I I \right\}$.

⁹When thinking about the discount rates firms apply to depreciation tax shields, this assumption feels appropriate. In general, our estimated λ spread will also include discount rate differences.

where z includes nontaxable states and possibly myopia, as in (A.7) and q now satisfies the general version of (A.11):

$$q_t = \sum_{s=t}^{\infty} \frac{1}{\Pi_{v=t}^s (1 + r_v + \delta)} [(1 + \lambda_s)(1 - \tau_s)\pi_{K,s} - \psi_{K,s}]. \quad (\text{A.17})$$

The only difference between (A.11) and (A.17) is that increasing capital leads to higher future after-tax profits, which relax future financial constraints.

We consider comparative statics with respect to θ , z_t^0 , λ , and γ . Except for λ , none of these terms directly affect q . They only affect q through investment's effect on future capital. We assume this latter effect is negligible. While nontrivial, this assumption is justified for two reasons. First, while the policies we study have a substantial temporary effect on investment, the change in investment is small relative to the existing capital stock. Thus, the long run marginal product of capital, which q measures, is likely unaffected.¹⁰ The second reason is that nearly all empirical studies of investment incentives assume that production exhibits constant returns to scale and linear homogeneity in adjustment costs, which leads to constant q as a function of capital.¹¹

Given the assumption that $\partial q / \partial \theta = 0$, our testable hypotheses build on the comparative static with respect to the bonus parameter θ :

$$\frac{\partial I}{\partial \theta} = \frac{(1 + \lambda)p_I [\gamma(\tau - z_t^0(\beta)) + (1 - \gamma)\beta\phi(\tau - z_t^0(1))]}{\psi_{II}} > 0. \quad (\text{A.18})$$

Bonus depreciation increases the present value of deductions, reducing the price of investment. Thus bonus depreciation should increase investment. Alternatively, we could study the effect of a general increase in z . The comparative static here is

$$\frac{\partial I}{\partial z} = \frac{(1 + \lambda)p_I}{\psi_{II}} > 0, \quad (\text{A.19})$$

¹⁰This is the assumption House and Shapiro (2008) make to replace short run approximations to capital and q with their steady state values. (See p. 740.)

¹¹Bond and Van Reenen (2007) survey the investment literature and argue that “conclusive evidence that linear homogeneity should be abandoned in the investment literature has not yet been presented.” This is because the assumption has both theoretical appeal and fits with evidence that changes in firm size are hard to predict (implying that firms do not have a sharp, optimal firm size).

which yields a useful equivalence between the depreciation elasticity, the price elasticity and the interest rate elasticity. In particular, $\varepsilon_{I,1-z} = \varepsilon_{I,p_I} \leq \varepsilon_{I,1+\lambda}$, where $\varepsilon_{I,x} = (\partial I / \partial x)(x/I)$ and the last inequality reflects the fact that $\partial q / \partial \lambda \geq 0$. We begin our empirical analysis by estimating different versions of (A.19), enabling easier comparisons to past work.

Hypothesis one concerns the differential effect of bonus depreciation on long and short duration industries. Long duration industries will have more delayed baseline deduction schedules and hence lower z_t^0 . The hypothesis thus derives from the cross partial of (A.18) with respect to z_t^0 :

$$\frac{\partial^2 I}{\partial \theta \partial z_t^0} = -\frac{(1+\lambda)p_I[\gamma + (1-\gamma)\beta\phi]}{\psi_{II}} < 0. \quad (\text{A.20})$$

Bonus results in relatively more acceleration for long lived items and so the investment response should be greater for these items. Note this is *not* a statement about the relative price elasticities for goods of different durations, which depend on the curvature of production and adjustment cost functions. Rather, it is a statement that bonus mechanically leads to larger price reductions for long duration items, even holding underlying technologies constant.

Hypothesis two concerns the differential effect of bonus depreciation for constrained and unconstrained firms. This depends on the cross partial of (A.18) with respect to λ :

$$\frac{\partial^2 I}{\partial \theta \partial \lambda} = \frac{\gamma(\tau - z_t^0(\beta)) + (1-\gamma)\beta\phi(\tau - z_t^0(1))}{\psi_{II}} p_I > 0. \quad (\text{A.21})$$

For constrained firms (i.e., when $\lambda > 0$), bonus both reduces the price of investment goods and relaxes the borrowing constraint. This is true even if the investment-cash flow sensitivity is zero, that is, if cash flow does not affect the marginal external finance cost, λ . The logic is similar to the foregoing logic about long versus short duration goods. The effective price cut due to bonus is larger for constrained firms, even if the cost of borrowing does not change. Under fairly general conditions therefore, financial constraints tend to amplify the effects of bonus.

Hypothesis three concerns the differential effect of bonus by taxable status. We can

compare the elasticities for γ equal to zero and γ equal to one:

$$\left. \frac{\partial I}{\partial \theta} \right|_{\gamma=1} - \left. \frac{\partial I}{\partial \theta} \right|_{\gamma=0} = (1 + \lambda) p_I \frac{(\tau - z_t^0(\beta)) - \beta \phi(\tau - z_t^0(1))}{\psi_{II}} > 0 \quad (\text{A.22})$$

Because nontaxable firms must wait to take bonus deductions, bonus is less valuable to them. This might be due to neoclassical reasons. Namely, taking into account a possibly long delay and applying a reasonable discount rate might lead the response for nontaxable firms to be quite low, even without myopia (i.e., with $\beta = 1$). We use the empirical distribution of loss transition probabilities to calibrate ϕ in the model and ask whether the results still require $\beta < 1$.

A.1.3 Empirical Moments for Calibration

We perform a calibration exercise to distinguish between models, based on their predictions about the external finance wedge, λ , and the discount rate applied to future flows, β . This exercise requires comparing estimates across subgroups. For this comparison to be useful, we need to make certain homogeneity assumptions about technologies across these groups. In particular, we want the curvature of adjustment costs to be equal across groups.

One way to satisfy this requirement is to make second derivatives effectively constant across groups. We make a weaker assumption, based on the common quadratic form used elsewhere in the literature. One feature of relying on this assumption is that nearly all other empirical studies of investment do so as well. Specifically, we write the adjustment cost function as

$$\psi(I, K) = \frac{\alpha}{2} [\log(I) - \log(\delta K)]^2 p_I I, \quad (\text{A.23})$$

so that adjustment costs are increasing quadratically as investment deviates from the replacement rate. As long as α is constant and average investment equals δK across groups, then the following results will hold.¹²

The first empirical moment we use compares the estimated response with respect to

¹²With our functional form for adjustment costs, we have $I\psi_{II} = \alpha p_I(1 + \log(I/\delta K))$, which is equal across groups under these assumptions.

bonus for constrained and unconstrained firms. Define the semi-elasticity of investment with respect to θ as $\varepsilon_{I,\theta} \equiv (\partial I / \partial \theta)(1/I)$, where $\partial I / \partial \theta$ is defined in (A.18). Assuming constrained firms face shadow price λ_C and unconstrained firms face shadow price λ_U , we take the ratio of semi-elasticities:

$$\frac{\varepsilon_{I,\theta}^C}{\varepsilon_{I,\theta}^U} \equiv m_1 = \frac{1 + \lambda_C}{1 + \lambda_U} = 1 + \frac{\Delta\lambda}{1 + \lambda_U}. \quad (\text{A.24})$$

We estimate m_1 and solve (A.24) for $\Delta\lambda / (1 + \lambda_U)$, which can be viewed as an implied credit spread. Our empirical analysis estimates the semi-elasticity with respect to z , rather than θ . Because z is linear in θ (see (A.6)), the ratio of z semi-elasticities equals the ratio of θ semi-elasticities.

We define a second empirical moment analogously by comparing taxable and nontaxable firms:

$$\frac{\varepsilon_{I,\theta}^{\gamma=0}}{\varepsilon_{I,\theta}^{\gamma=1}} \equiv m_2 = \beta\phi \frac{\tau - z_t^0(1)}{\tau - z_t^0(\beta)} \quad (\text{A.25})$$

Note the external finance wedge falls out of this expression. This is true as long as average shadow costs are the same across taxable and nontaxable groups.¹³ Under a constant τ assumption, we can drop tax rates from this formula, which we do in Section 1.5.3. We estimate m_2 and calibrate ϕ in order to estimate β .

A.2 Legislative Background

This appendix describes legislation affecting the bonus and Section 179 depreciation provisions studied in this paper.

Economic Recovery Tax Act of 1981

The act set the Section 179 allowance at \$5,000 and established a timetable for gradually increasing the allowance to \$10,000 by 1986. Few firms took advantage of the allowance

¹³We can relax this assumption, since we expect nontaxable firms to be more constrained on average. Alternatively, we can narrow our taxable/nontaxable comparison to groups that differ only by how likely it is for the next dollar of investment to affect this year's taxes. We pursue this latter approach and use the stock of alternative tax shields to sort firms.

initially. Some attributed the low response to limitations on the use of the investment tax credit. A business taxpayer could claim the credit only for the portion of an eligible asset's cost that was not expensed; so the full credit could be used only if the company claimed no expensing allowance. For many firms, the tax savings from the credit alone outweighed the tax savings from combining the credit with the allowance.¹⁴

Depreciation Policies Affected – Section 179

Date Signed – August 13, 1981

Bill Number – H.R. 4242

Deficit Reduction Act of 1984

The act postponed from 1986 to 1990 the scheduled increase in the Section 179 allowance to \$10,000. Use of the allowance rose markedly following the repeal of the investment tax credit by the Tax Reform Act of 1986.

Depreciation Policies Affected – Section 179

Date Signed – July 18, 1984

Bill Number – H.R. 4170

Omnibus Budget Reconciliation Act of 1993

The act increased the Section 179 allowance from \$10,000 to \$17,500, as of January 1, 1993.

Depreciation Policies Affected – Section 179

Date Introduced – May 25, 1993

Date of First Passage Vote – May 27, 1993

Date Signed – August 10, 1993

Bill Number – H.R. 2264

¹⁴Source: http://www.section179.org/stimulus_acts.html

Small Business Job Protection Act of 1996

The act increased the Section 179 allowance and established scheduled annual (with one exception) increases over six years. Specifically, the act raised the maximum allowance to \$18,000 in 1997, \$18,500 in 1998, \$19,000 in 1999, \$20,000 in 2000, \$24,000 in 2001 and 2002, and \$25,000 in 2003 and thereafter.

Depreciation Policies Affected – Section 179

Date Introduced – May 14, 1996

Date of First Passage Vote – May 22, 1996

Date Signed – August 20, 1996

Bill Number – H.R. 3448

Job Creation and Worker Assistance Act of 2002

The act created the first bonus depreciation allowance, equal to 30 percent of the adjusted basis of new qualified property acquired after September 11, 2001, and placed in service no later than December 31, 2004. A one-year extension of the placed-in-service deadline was available for certain property with a MACRS recovery period of 10 or more years and for transportation equipment.

Depreciation Policies Affected – Bonus Depreciation

Date Introduced – October 11, 2001

Date of First Passage Vote – October 24, 2001

Date Signed – March 9, 2002

Bill Number – H.R. 3090

Jobs and Growth Tax Relief Reconciliation Act of 2003

The act (JGTRRA) raised the bonus allowance to 50 percent for qualified property acquired after May 5, 2003, and placed in service before January 1, 2005. The act raised the Section 179 allowance to \$100,000 (as of May 6, 2003), set it to stay at that amount in 2004 and 2005, and then reset in 2006 and beyond at its level before JGTRRA (\$25,000). JGTRRA also raised the phase out threshold to \$400,000 from May 2003 to the end of 2005, indexed the regular allowance and the threshold for inflation in 2004 and 2005, and added off-the-shelf software for business use to the list of depreciable assets eligible for expensing in the same period.

The American Jobs Creation Act of 2004 extended the Section 179 changes made by JGTRRA through the end of 2007. The Tax Increase Prevention and Reconciliation Act of 2005 extended the changes in the allowance under JGTRRA through 2009.

Depreciation Policies Affected – Bonus Depreciation and Section 179

Date Introduced – February 27, 2003

Date of First Passage Vote – May 9, 2003

Date Signed – May 28, 2003

Bill Number – H.R. 2

U.S. Troop Readiness, Veterans' Care, Katrina Recovery, and Iraq Appropriations Act of 2007

Congress extended the changes in the allowance made by JGTRRA through 2010, raised the maximum allowance to \$125,000 and the phaseout threshold to \$500,000 for 2007 to 2010, and indexed both amounts for inflation in that period.

Depreciation Policies Affected – Section 179

Date Introduced – May 8, 2007

Date of First Passage Vote – May 10, 2007

Date Signed – May 25, 2007

Bill Number – H.R. 2206

Economic Stimulus Act of 2008

The act provided for 50 percent bonus depreciation. To claim the allowance, a taxpayer had to acquire qualified property after December 31, 2007 and place it in service before January 1, 2009. The previous \$125,000 limit on the Section 179 allowance was increased to \$250,000, and the \$500,000 limit on the total amount of equipment purchased became \$800,000.

Depreciation Policies Affected – Bonus Depreciation and Section 179

Date Introduced – January 28, 2008

Date of First Passage Vote – January 29, 2008

Date Signed – February 13, 2008

Bill Number – H.R. 5140

American Recovery and Reinvestment Act of 2009

The act extended the deadlines by one year, to the end of 2009, for the 50 percent bonus depreciation allowance.

Depreciation Policies Affected – Bonus Depreciation

Date Introduced – January 26, 2009

Date of First Passage Vote – January 28, 2009

Date Signed – February 17, 2009

Bill Number – H.R. 1

Small Business Jobs Act of 2010

The act extended the 50 percent bonus depreciation to qualifying property purchased and placed in service during the 2010 tax year. The act increased the amount a business could expense under Section 179 from \$250,000 to \$500,000 of qualified capital expenditures. These deductions were subject to a phase-out for expenditures exceeding \$2,000,000. The provision covered tax years for 2010 and 2011.

Depreciation Policies Affected – Bonus Depreciation and Section 179

Date Introduced – May 13, 2010

Date of First Passage Vote – June 17, 2010

Date Signed – September 27, 2010

Bill Number – H.R. 5297

Tax Relief, Unemployment Compensation Reauthorization, and Job Creation Act of 2010

The bonus depreciation allowance increased to 100 percent for qualified property acquired after September 8, 2010, and placed in service before January 1, 2012. The act also established a 50 percent allowance for property acquired and placed in service in 2012.

Depreciation Policies Affected – Bonus Depreciation

Date Introduced – March 16, 2010

Date Signed – September 27, 2010

Bill Number – H.R. 5297

Table A.1.1: *Section 179 and Bonus Depreciation Policy Changes*

Year	S179 Max Value	S179 Phase-out Region	Bonus
1993-96	\$17,500	\$200,000-\$217,500	
1997	\$18,000	\$200,000-\$218,000	
1998	\$18,500	\$200,000-\$218,500	
1999	\$19,000	\$200,000-\$219,000	
2000	\$20,000	\$200,000-\$220,000	
2001-02	\$24,000	\$200,000-\$224,000	30% Tax years ending after 9/10/01
2003	\$100,000	\$400,000-\$500,000	50% Tax years ending after 5/3/03
2004	\$102,000	\$410,000-\$512,000	50%
2005	\$105,000	\$420,000-\$525,000	
2006	\$108,000	\$430,000-\$538,000	
2007	\$125,000	\$500,000-\$625,000	
2008-09	\$250,000	\$800,000-\$1,050,000	50% Tax years ending after 12/31/07
2010-11	\$500,000	\$2,000,000-\$2,500,000	100% Tax years ending after 9/8/10

a. 2008 was retroactive.

Table A.1.2: Past User Cost Estimates

paper	equation	β_1 (SE)	estimation details	data	table / page cite
Cummins, Hassett, and Hubbard (1994)	$\frac{I}{K} = \beta_0 + \beta_1 Q$	0.083(0.006)	first-differences; firm and year FEs; robust SE; all-years	US public firm panel (Compustat), 1953-88	Table 4 (OLS, all years) / p. 28
			first-differences; robust SE; 1962 (major tax reform)	US public firm panel (Compustat), 1953-88	Table 4 (OLS, 1962) / p. 28
			first-differences; robust SE; 1972 (major tax reform)	US public firm panel (Compustat), 1953-88	Table 4 (OLS, 1972) / p. 28
			first-differences; robust SE; 1981 (major tax reform)	US public firm panel (Compustat), 1953-88	Table 4 (OLS, 1981) / p. 28
			first-differences; robust SE; 1986 (major tax reform)	US public firm panel (Compustat), 1953-88	Table 4 (OLS, 1986) / p. 28
Cummins, Hassett, and Hubbard (1996)	$\frac{I}{K} = \beta_0 + \beta_1 Q$	0.647(0.238)	difference observed and forecasted variables; forecasting based on lagged $\frac{I}{K}$, lagged $\frac{CF}{K}$, time-trend, and firm FE; robust SE; AUS 1988	Int'l public firm panel (Global Vantage), 1982-92	Table 6 (AUS 1988, top) / p. 254
			same as above; BEL 1990	Int'l public firm panel (Global Vantage), 1982-92	Table 6 (BEL 1990, top) / p. 254
			same as above; CAN 1988	Int'l public firm panel (Global Vantage), 1982-92	Table 6 (CAN 1988, top) / p. 254

Table A.1.2: *(continued)*

0.867(0.458)	same as above; DNK 1988	Int'l public firm	Table 6 (DNK 1990, panel (Global Van-top) / p. 254 tage), 1982-92
0.756(0.286)	same as above; FRA 1990	Int'l public firm	Table 6 (FRA 1990, panel (Global Van-top) / p. 254 tage), 1982-92
0.938(0.242)	same as above; GER 1990	Int'l public firm	Table 6 (GER 1990, panel (Global Van-top) / p. 254 tage), 1982-92
0.663(0.237)	same as above; ITA 1992	Int'l public firm	Table 6 (ITA 1992, panel (Global Van-top) / p. 254 tage), 1982-92
0.893(0.219)	same as above; JPN 1989	Int'l public firm	Table 6 (JPN 1989, panel (Global Van-top) / p. 254 tage), 1982-92
0.423(0.340)	same as above; NLD 1989	Int'l public firm	Table 6 (NLD 1989, panel (Global Van-top) / p. 254 tage), 1982-92
1.373(0.528)	same as above; NOR 1992	Int'l public firm	Table 6 (NOR 1992, panel (Global Van-top) / p. 254 tage), 1982-92

Table A.1.2: *(continued)*

1.485(1.378)	same as above; SPN 1989	Int'l public firm panel (Global Vantage), 1982-92	Table 6 (SPN 1989, top) / p. 254
0.641(0.241)	same as above; SWE 1990	Int'l public firm panel (Global Vantage), 1982-92	Table 6 (SWE 1990, top) / p. 254
0.644(0.198)	same as above; UK 1991	Int'l public firm panel (Global Vantage), 1982-92	Table 6 (UK 1991, top) / p. 254
0.603(0.086)	same as above; USA 1987	Int'l public firm panel (Global Vantage), 1982-92	Table 6 (USA 1987, top) / p. 254
Desai and Goolsbee (2004)	year and firm FEs; SE clustered at firm-level	U.S. public firm panel (Computstat), 1962-03	Table 8 (baseline) / p. 314
Edgerton (2010)	year and firm FEs; SE clustered at firm-level; includes dummy and interaction for non-taxable firms	US public firm panel (Computstat), 1967-05	Table 3 (2) / p. 945

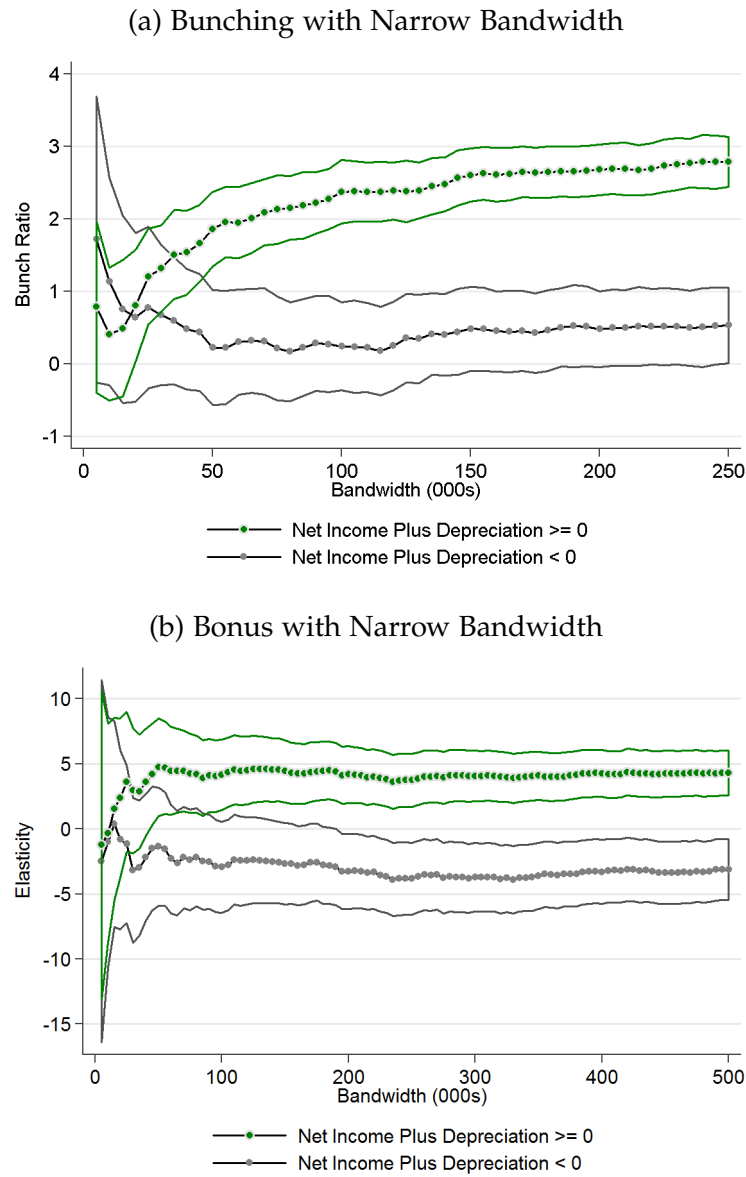


Figure A.1.1: Investment Behavior and Tax Incentives: Narrow Bandwidth

Notes: These figures replicate the taxable position splits in the bunch and bonus settings, while restricting the sample to within a narrow bandwidth of the tax status threshold. Panel (a) replicates the analysis in panel (a) of Figure 1.4, which compares bunching behavior for taxable and nontaxable firms. Panel (b) replicates the regression in column (1) of table 1.5, which estimates separate coefficients with respect to bonus incentives for taxable and nontaxable firms.

Appendix B

Appendix to Chapter 2

Micro-foundation of owner-occupancy utility. We present a moral hazard framework in which ownership utility matches the specification of (2.2). Our framework follows the spirit of Henderson and Ioannides (1983)'s treatment of tenure choice, in which maintenance frictions lead some residents to own instead of rent.

Residents derive utility from the particular way their house is “customized”: e.g. the color of the walls, the way the lawn is maintained, et cetera. The set of possible customizations is \mathcal{K} . Resident i 's utility from housing is $v(\sum_{k \in \mathcal{K}} a_{i,k} h_k)$, where $a_{i,k} > 0$ is his preference for k and h_k is the quantity of housing customized that way. Individual customization choices are not contractible, but the right to customize one's house is. If a landlord retains the customization rights, then the tenant cannot customize the house, and the null customization $k = 0$ occurs for which $a_{i,0} = 1$ for all residents i . If the tenant holds these rights, he may choose any $k \in \mathcal{K} \setminus \{0\}$.

Moral hazard arises due to a doomsday customization $k = d$. This customization incurs a cost $\eta(h_d)$ to the owner of the house. All residents prefer this customization to all others: $d = \arg \max_{k \in \mathcal{K}} a_{i,k}$. However, the costs of d outweigh the benefits: for all i and h ,

$$v(a_{i,d}h) < \eta(h).$$

The doomsday customization represents the proclivity of residents to damage a house when

they do not bear the costs of doing so.

This inequality prevents landlords from ever selling customization rights to tenants. Suppose the landlord sells the rights. Then the tenant chooses his preferred customization, without taking into account the resultant costs, which the landlord bears. The tenant therefore chooses $k = d$. Knowing this, the landlord demands at least $\eta(h)$ for the customization rights. But the most the tenant is willing to pay is $v(a_{i,d}h) - v(h)$, which is less than $\eta(h)$. Therefore they agree not to trade. The landlord keeps the rights, and $k = 0$. The utility from renting is $v(h)$ because $a_{i,0} = 1$.

An owner-occupant chooses the customization, but also bears the costs if he chooses $k = d$. Let $k(i)$ denote the solution to his optimization problem $\max_{k \in \mathcal{K} \setminus \{0\}} v(a_{i,k}h) - \eta(h)\mathbf{1}_{k=d}$. Due to the costliness of the doomsday customization, the resident never chooses it: $k(i) \neq d$. Indeed, if k' is any other customization, then $v(a_{i,k'}h) > v(a_{i,d}h) - \eta(h)$ due to the above inequality. We define $a_i \equiv a_{i,k(i)}$. The utility from owning is $v(a_ih)$. This form corresponds exactly to (2.2).

Proof of Lemma 1. First we prove that construction occurs in each period. Construction occurs at time 0 because the housing stock starts at 0, and the housing demand equation (2.9) is positive. For a contradiction, let $t_1 > 0$ denote the first period in which construction does not occur. Let $t_2 > t_1$ denote the next time construction occurs (t_2 may be infinite).

We now claim that $r_t^h > r_{t_1-1}^h$ for $t_1 \leq t < t_2$. Along the trend growth path, $x = 0$, so no uncertainty exists and by (2.7), a resident rents if and only if $a_i < 1$. Because F_a has full support on \mathbb{R}^+ , some residents must rent. Landlords hence exist in equilibrium, and their arbitrage equation $p_t^h = r_t^h + \beta p_{t+1}^h$ holds. Aggregate housing demand resulting from the first-order condition (2.6) is

$$D_t^h(r_t^h) = N_t \left(\int_0^1 (v')^{-1}(r_t^h) dF_a + \int_1^\infty (v')^{-1}(r_t^h/a_i) / a_i dF_a \right). \quad (\text{A1})$$

By assumption, the housing stock and hence housing demand is the same for $t_1 - 1 \leq t < t_2$. Equation (A1) decreases in r_t^h because $v'' < 0$. Because $x = 0$ and $g > 0$, N_t increases with t . The left side of (A1) stays constant for $t_1 - 1 \leq t < t_2$ while N_t increases. Therefore, r_t

increases for $t_1 - 1 \leq t < t_2$.

Because construction occurs at $t_1 - 1$, we have $p_{t_1-1}^h = p_{t_1-1}^l + K$, which results from zero homebuilder profits. Zero construction at t_1 can only occur when $p_{t_1}^h \leq p_{t_1}^l + K$, from homebuilder profit maximization. The landlord and landowner arbitrage equations at $t_1 - 1$ deliver $r_{t_1-1}^h \geq r_{t_1-1}^l + (1 - \beta)K$. The quantity of undeveloped land stays constant for $t_1 - 1 \leq t < t_2$ and hence r_t^l does as well, because firm land demand D^l does not change over time. Therefore $r_t^h > r_t^l + (1 - \beta)K$ for $t_1 \leq t < t_2$. Then

$$\begin{aligned} p_{t_1}^h &= \sum_{t_1 \leq t < t_2} \beta^{t-t_1} r_t^h + \beta^{t_2-t_1} p_{t_2}^h \\ &> \sum_{t_1 \leq t < t_2} \beta^{t-t_1} (r_t^l + (1 - \beta)K) + \beta^{t_2-t_1} (p_{t_2}^l + K) \\ &= p_{t_1}^l + K, \end{aligned}$$

which contradicts the zero construction inequality $p_{t_1}^h \leq p_{t_1}^l + K$. This contradiction proves that construction occurs at all times t .

We now show that rents r_t^h increase over time. Because construction occurs at all t , $p_t^h = p_t^l + K$ for all t . Undeveloped land must always exist because perpetual construction occurs. Therefore, landowners are indifferent between holding land until tomorrow or selling it, so $p_t^l = r_t^l + \beta p_{t+1}^l$. Together with the landlord arbitrage equation, this equation gives $r_t^h = p_t^h - \beta p_{t+1}^h = p_t^l + K - \beta(p_{t+1}^l + K) = r_t^l + (1 - \beta)K$. Equilibrium rents are determined by $S - D^l(r_t^h - (1 - \beta)K) = D_t^h(r_t^h)$, where housing demand comes from (A1). The left side increases in r_t^h , whereas the right side decreases. N_t increases over time, which shifts up D_t^h . Therefore r_t^h increases as well.

Finally, we can show directly that the supply elasticity decreases over time. The elasticity by definition is

$$\epsilon_t^S = \frac{r_t^h (D^l)'(r_t^h - (1 - \beta)K)}{S - D^l(r_t^h - (1 - \beta)K)} = \frac{r_t^h}{r_t^h - (1 - \beta)K} \frac{D^l(r_t^h)}{D_t^h(r_t^h)} \frac{r_t^l (D^l)'(r_t^l)}{D^l(r_t^l)} = \frac{r_t^h}{r_t^h - (1 - \beta)K} \left(\frac{S}{H_t} - 1 \right) \epsilon^l,$$

which coincides with (2.4). We have shown directly that H_t and r_t^h increase over time. Therefore, when ϵ^l is constant, ϵ_t^S decreases over time.

Proof of Proposition 2. We use (2.5) to write $p_0^h = r_0^h + \beta \tilde{\mathbf{E}} p_1^h$. Let $\partial/\partial x$ denote the partial derivative in which N_0 stays constant. Then $\partial p_0^h/\partial x = \partial r_0^h/\partial x + \partial \beta \tilde{\mathbf{E}} p_1^h/\partial x$. We calculate $\partial r_0^h/\partial x$ by differentiating (2.8) at $x = 0$. Let $d(\cdot) = (v')^{-1}(\cdot)$, and let $b_i = 1 + \beta(\tilde{\mathbf{E}} p_1^h - \mathbf{E}_i p_1^h)/r_0^h$. Note that when $x = 0$, $b_i = 1$ for all i . Then

$$\begin{aligned} -(D^l)' \frac{\partial r_0^h}{\partial x} &= N_0 \int_M \int_0^1 d'(r_0^h) \frac{\partial r_0^h}{\partial x} dF_a dF_\mu + N_0 \int_M d(r_0^h) \frac{\partial b_i}{\partial x} dF_\mu \\ &+ N_0 \int_M \int_1^\infty a_i^{-2} d'(r_0^h/a_i) \left(\frac{\partial r_0^h}{\partial x} + \frac{\partial \beta \tilde{\mathbf{E}} p_1^h}{\partial x} - \frac{\partial \beta \mathbf{E}_i p_1^h}{\partial x} \right) dF_a dF_\mu - N_0 \int_M d(r_0^h) \frac{\partial b_i}{\partial x} dF_\mu. \end{aligned}$$

The extensive margins terms for the rental and owner-occupied populations cancel. We simplify this equation to

$$\frac{\partial r_0^h}{\partial x} = - \frac{N_0 \int_M \int_1^\infty a_i^{-2} d'(r_0^h/a_i) \left(\partial \beta \tilde{\mathbf{E}} p_1^h/\partial x - \partial \beta \mathbf{E}_i p_1^h/\partial x \right) dF_a dF_\mu}{(D^l)' + N_0 \int_M \int_0^1 d'(r_0^h) dF_a dF_\mu + N_0 \int_M \int_1^\infty a_i^{-2} d'(r_0^h/a_i) dF_a dF_\mu}.$$

The proposition assumes a constant elasticity of housing demand ϵ^D . This property occurs when individual demand $d(\cdot)$ displays the same constant elasticity. Indeed, from (A1), the elasticity of housing demand when $x = 0$ is

$$\epsilon^D = - \frac{\int_0^1 r d'(r) dF_a + \int_1^\infty r a_i^{-2} d'(r/a_i) dF_a}{\int_0^1 d(r) dF_a + \int_1^\infty a_i^{-1} d(r/a_i) dF_a},$$

which holds when $r d'(r)/d(r) = -\epsilon^D$ for all r . We can therefore rewrite $\partial r_0^h/\partial x$ as

$$\frac{\partial r_0^h}{\partial x} = - \frac{\epsilon^D N_0 \int_M \int_1^\infty a_i^{-1} d(r_0^h/a_i) \left(\partial \beta \tilde{\mathbf{E}} p_1^h/\partial x - \partial \beta \mathbf{E}_i p_1^h/\partial x \right) dF_a dF_\mu}{r_0^h (D^l)' + \epsilon^D N_0 \int_M \int_0^1 d(r_0^h) dF_a dF_\mu + \epsilon^D N_0 \int_M \int_1^\infty a_i^{-1} d(r_0^h/a_i) dF_a dF_\mu}.$$

Because F_a and F_μ are independent, we can write

$$\begin{aligned} \int_M \int_1^\infty a_i^{-1} d(r_0^h/a_i) \mathbf{E}_i p_1^h dF_a dF_\mu &= \int_1^\infty a_i^{-1} d(r_0^h/a_i) dF_a \int_M \mathbf{E}_i p_1^h dF_\mu \\ &= \int_1^\infty a_i^{-1} d(r_0^h/a_i) dF_a \bar{\mathbf{E}} p_1^h, \end{aligned}$$

where $\bar{\mathbf{E}} p_1^h \equiv \int_M \mathbf{E}_i p_1^h dF_\mu$ is the average belief about p_1^h . Recall from (A1) that $(h_{i,0}^{rent})^* = d(r_0^h)$ if $a_i < 1$ (and 0 otherwise) and $(h_{i,0}^{own})^* = d(r_0^h/a_i)/a_i$ if $a_i \geq 1$ (and 0 otherwise).

The share of housing that is owner-occupied is $\chi = \int_1^\infty a_i^{-1} d(r_0^h/a_i) dF_a / (\int_0^1 d(r_0^h) dF_a +$

$\int_1^\infty a_i^{-1} d(r_0^h/a_i) dF_a$). We can therefore divide through the equation for $\partial r_0^h/\partial x$ by the total housing stock to get

$$\frac{\partial r_0^h}{\partial x} = -\frac{\epsilon^D \chi \left(\partial \beta \tilde{\mathbf{E}} p_1^h / \partial x - \partial \beta \bar{\mathbf{E}} p_1^h / \partial x \right)}{\epsilon_0^S + \epsilon^D}.$$

Substituting into $\partial p_0^h/\partial x = \partial r_0^h/\partial x + \partial \beta \tilde{\mathbf{E}} p_1^h/\partial x$ yields (2.10) of the proposition.

Proof of Proposition 3. We will calculate the effect of the shock z_t on r_t^h by differentiating the equation $S - D^l(r_t^h - (1 - \beta)K) = D_t^h(r_t^h)$ with respect to x at $x = 0$, where $D_t^h(r_t^h)$ is given by (A1). This derivative is valid if and only if this equilibrium condition holds for x around 0. The condition holds as long as construction occurs at t . Our first task is thus proving the existence of an open set $I \in \mathbb{R}$ such that $0 \in I$ and for $x \in I$, construction occurs for all t .

As in the proof of Lemma 1, we can prove that construction must occur at t_1 if, conditional on the absence of construction at t_1 , $r_t > r_{t_1-1}^h$ for $t_1 \leq t < t_2$ where t_2 is the next time construction occurs. The key step in that proof was that N_t increases with t . We define an open set I_1 containing 0 such that N_t still increases in t for $x \in I_1$. Because M is uniformly bounded, there exist μ^{\min} and μ^{\max} such that $\mu^{\min} \leq \mu' \leq \mu^{\max}$ for all μ' that are coordinates of vectors in M . Recall that $N_{t+1}/N_t = e^{g+(\mu_{t+1}-\mu_t)x}$. Because $g > 0$, the set $I_1 = (-g/(\mu^{\max} - \mu^{\min}), g/(\mu^{\max} - \mu^{\min}))$ is open. For any $x \in I_1$, $N_{t+1}/N_t > 1$. With this result, the proof of this increasing rent condition matches verbatim the proof given in the proof of Lemma 1 when $t_1 > 1$. When $t_1 = 1$, $D_{t_1-1}^h$ is no longer given by (A1) but instead by (2.9).

The only new fact we must show therefore is that if construction fails to occur at $t = 1$, then $r_0^h < r_1^h$. To do this, we first show that $\tilde{\mathbf{E}} p_1^h - \mathbf{E}_i p_1^h = O(x)$ as $x \rightarrow 0$ for all i . We have $p_1^h = \sum_{t=1}^\infty \beta^{t-1} r_t^h$. All residents agree on H_0 and N_0 because they are observable at $t = 0$. Let t_2 be the next time construction occurs given H_0 . Once it occurs it will occur afterward forever due to the arguments in the proof of Lemma 1. In principle residents could disagree about t_2 , but we will now show that for x small enough they do not. While construction does not occur, rents are determined by $H_0 = N_t D_t^h(r_t^h)$ and $S - H_0 = D^l(r_t^l)$. Because N_t increases over time, r_t^h must as well. When construction

occurs next period but not today at t , $p_t^h < p_t^l + K$ while $p_{t+1}^h = p_{t+1}^l + K$, so using the landlord and landowner arbitrage equations defined in the proof of Lemma 1, we find that $(D_t^h)^{-1}(H_0/N_t) < (D^l)^{-1}(S - H_0) + (1 - \beta)K$ while construction fails to occur. The first time construction does occur, $t = t_2$, is defined as the lowest value of t for which this inequality fails to hold. Because we are in discrete time, and because the relationships $N_t = N_0 e^{g^t + (\mu_t - 1)x}$ and $\mu^{\min} \leq \mu_t \leq \mu^{\max}$ hold, there exists an open $I_2 \ni 0$ such that when $x \in I_2$, t_2 is the same for all realizations of $\mu \in M$. For $1 \leq t < t_2$, r_t^h is the solution to $H_0 = N_t D_t^h(r_t^h)$, and for $t \geq t_2$, r_t^h solves $S - D^l(r_t^h - (1 - \beta)K) = N_t D_t^h(r_t^h)$. In each case, because $N_t = N_0 e^{g^t + (\mu_t - 1)x}$, the resulting r_t^h is a differentiable function of x for any value of μ_t and is the same at $x = 0$ for any value of μ_t . Therefore, $\tilde{E}r_t^h - E_i r_t^h = O(x)$ as $x \rightarrow 0$ for all i , and the same then holds for p_1^h .

We now return to showing that if construction fails to occur at $t = 1$, then $r_0^h < r_1^h$. Using (2.9), we write $D_0^h(r_0^h) = N_0 f_0(r_0^h)$, and using (A1), we write $D_1^h(r_1^h) = N_1 f_1(r_1^h)$. Without construction at $t = 1$, we have $N_0 f_0(r_0^h) = N_1 f_1(r_1^h)$. Note from (2.9) and (A1) that $f_0 = f_1 + O(x)$ as $x \rightarrow 0$; this fact follows because $\tilde{E}p_1^h - E_i p_1^h = O(x)$ as $x \rightarrow 0$ for all i . Using $N_1 = N_0 e^{g + (\mu_1 - 1)x}$, we can conclude that $e^{g + (\mu_1 - 1)x} f_1(r_1^h) = f_1(r_0^h) + O(x)$ as $x \rightarrow 0$. Because $e^{g + (\mu_1 - 1)x} > 1$ as $x \rightarrow 0$ and f_1 is decreasing, there exists an open $I_3 \ni 0$ such that for $x \in I_3$, $r_1^h > r_1^0$. This inequality is what we needed to show to prove that construction occurs at time 1, which is all that remained to prove that construction always occurs. We set $I = I_1 \cap I_2 \cap I_3$.

All of that proved that for $t > 0$, the effect of the shock z_t on r_t^h results from differentiating the equation $S - D^l(r_t^h - (1 - \beta)K) = D_t^h(r_t^h)$ with respect to x at $x = 0$. Doing so yields $-(D^l)' dr_t^h / dx = \mu_t D_t^h + (D_t^h)' dr_t^h / dx$, from which it follows that $dr_t^h / dx = -\mu_t D_t^h / ((D^l)' + (D_t^h)') = \mu_t r_t^h / (\epsilon_t^S + \epsilon^D)$. Similarly, the partial effect of the shock on current rents r_0^h , holding beliefs constant and letting N_0 change, is $\partial r_0^h / \partial x = r_0^h / (\epsilon_0^S + \epsilon^D)$. Putting together this partial effect with the one in Proposition 2 yields

$$\frac{dp_0^h}{dx} = \frac{r_0^h}{\epsilon_0^S + \epsilon^D} + \sum_{t=1}^{\infty} \left(\frac{\epsilon_0^S + (1 - \chi)\epsilon^D}{\epsilon_0^S + \epsilon^D} \tilde{\mu}_t + \frac{\chi\epsilon^S}{\epsilon_0^S + \epsilon^S} \bar{\mu}_t \right) \frac{\beta^t r_t^h}{\epsilon_t^S + \epsilon^S},$$

where $\tilde{\mu}_t$ is the most optimistic belief of μ_t and $\bar{\mu}_t$ is the average belief of μ_t . Because all residents agree that $\mu_0 = 1$, we may rewrite this expression as

$$\frac{dp_0^h}{dx} = \sum_{t=0}^{\infty} \left(\frac{\epsilon_0^S + (1-\chi)\epsilon^D}{\epsilon_0^S + \epsilon^D} \tilde{\mu}_t + \frac{\chi\epsilon^S}{\epsilon_0^S + \epsilon^S} \bar{\mu}_t \right) \frac{\beta^t r_t^h}{\epsilon_t^S + \epsilon^S}.$$

The text defines the mean persistence of the shock μ to be $\mu = \sum_{t=0}^{\infty} \mu_t \beta^t r_t^h (\epsilon_t^S + \epsilon^D)^{-1} / \sum_{t=0}^{\infty} \beta^t r_t^h (\epsilon_t^S + \epsilon^D)^{-1}$. We use this definition, and divide through by $p_0 = \sum_{t=0}^{\infty} \beta^t r_t^h$, which holds at $x = 0$, to derive

$$\begin{aligned} \frac{d \log p_0^h}{dx} &= \left(\sum_{t=0}^{\infty} \beta^t r_t^h \right)^{-1} \sum_{t=0}^{\infty} \left(\frac{\epsilon_0^S + (1-\chi)\epsilon^D}{\epsilon_0^S + \epsilon^D} \tilde{\mu} + \frac{\chi\epsilon^S}{\epsilon_0^S + \epsilon^S} \bar{\mu} \right) \frac{\beta^t r_t^h}{\epsilon_t^S + \epsilon^S} \\ &= \left(\frac{\epsilon_0^S + (1-\chi)\epsilon^D}{\epsilon_0^S + \epsilon^D} \tilde{\mu} + \frac{\chi\epsilon^S}{\epsilon_0^S + \epsilon^S} \bar{\mu} \right) \frac{1}{\tilde{\epsilon}^S + \epsilon^D}, \end{aligned}$$

where we have used the definition of the long-run supply elasticity $\tilde{\epsilon}^S$ given in the text. This equation for $d \log p_0^h / dx$ matches (2.12) in Proposition 3.

Proof of Implication 7. We demonstrate a limiting case in which $\epsilon_0^S = \infty$ while $\tilde{\epsilon}^S < \infty$. Let $D^l(r) = br^{-\epsilon^l}$ for some constant $b > 0$. Consider the limit as $b \rightarrow 0$. We know that $r_t^h \geq (1-\beta)K$ because $r_t^h = r_t^l + (1-\beta)K$ and $r_t^l \geq 0$. Define N^* to be the value of N_t that solves the equation $S = D_t^h((1-\beta)K)$, where D_t^h is given by (A1). For $N_t < N^*$, housing demand fails to exceed available land at the minimum rent, and there is no demand for land in the limit, so the market clearing rent must be $r_t^h = (1-\beta)K$ while $H_t < S$. By (2.4), $\epsilon_t^S = \infty$ in this case. But for $N_t > N^*$, demand exceeds supply at the minimum rent, so $r_t^h > (1-\beta)K$ and $H_t > 0$, leading to a finite elasticity. Since N_t grows at a constant rate g , for any $N_t < N^*$ we have $\epsilon_0^S = \infty$ but $\tilde{\epsilon}^S < \infty$.

Proof of Implication 8. Disagreement amplification Δ equals

$$\Delta = \frac{\epsilon_0^S + (1-\chi)\epsilon^D}{\epsilon_0^S + \epsilon^D} \frac{\tilde{\mu} - \bar{\mu}}{\tilde{\epsilon}^S + \epsilon^D}.$$

We calculate this difference from subtracting from (2.12) the counterfactual in which we substitute $\bar{\mu}$ for $\tilde{\mu}$. Define $\bar{N}_0^*(\chi)$ to be the value of development (which determines the supply elasticities; see above) that maximizes Δ . When $\chi = 1$, Δ is 0 in the limits as $\bar{N}_0 \rightarrow 0$

and $\bar{N}_0 \rightarrow \infty$, because $\tilde{\epsilon}^S = 0$ in the first case and $\epsilon_0^S = 0$ in the second. But $\Delta > 0$ for $\chi = 1$, so $0 < \bar{N}_0^*(1) < \infty$ by continuity. But $\bar{N}_0^*(\chi)$ is continuous in χ as long as it exists and is finite, so there must exist $\chi^* < 1$ such that for $\chi^* \leq \chi \leq 1$, $\bar{N}_0^*(\chi)$ exists and is finite.

Proof of Implication 9. When $\chi = 1$, the limit as $\bar{N}_0 \rightarrow \infty$ of $d \log p_0^h / dx$ is $\bar{\mu} / \epsilon^D$. For any $0 < \bar{N}_0 < \infty$, we can choose $\tilde{\mu}$ to be large enough so that the price change given by (2.12) is larger than $\bar{\mu} / \epsilon^D$, because this price change becomes arbitrarily large with $\tilde{\mu}$. By continuity, we can do the same for some $\chi < 1$.

Construction equation. By the definition of supply elasticity, the change in the log housing stock is $\epsilon_0^S d \log r_0^h / dx$. The total effect of the shock on rents combines the effect in the end of the proof of Proposition 2 and the direct effect of the shock on N_0 derived in the proof of Proposition 3. It is $dr_0^h / dx = r_0^h / (\epsilon_0^S + \epsilon^D) - \chi \epsilon^D (\partial \beta \tilde{\mathbf{E}} p_1^h / \partial x - \partial \beta \bar{\mathbf{E}} p_1^h / \partial x) / (\epsilon_0^S + \epsilon^D)$. We substitute in for the beliefs from the Proof of Proposition 3 and divide through by r_0^h , and then multiply by ϵ_0^S to get

$$\frac{d \log H_0}{dx} = \frac{\epsilon_0^S}{\epsilon_0^S + \epsilon^D} \left(1 - \frac{\chi \epsilon^D}{\tilde{\epsilon}^S + \epsilon^D} \rho (\tilde{\mu} - \bar{\mu}) \right), \quad (\text{A2})$$

where $\rho \equiv p_0^h / r_0^h$ is the price-rent ratio of the city before the shock at $x = 0$.